REPORT RESUMES

ED 003 883 48

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION--III. PROGRESS REPORT NO. 3.

BY- LANE, HARLAN AND OTHERS

MICHIGAN UNIV., ANN ARBOR,COLL.OF LIT.,SCI.,ARTS

REPORT NUMBER NDEA-VI-77-3 PUB DATE APR 63

REPORT NUMBER PR-3

CONTRACT OEC-SAE-9265

EDRS PRICE MF-$0.36 HC-$10.00 250P.

DESCRIPTORS- *SPEECH SKILLS, SPEECH, STRUCTURAL ANALYSIS, *SPEECH INSTRUCTION, STIMULUS DEVICES, REINFORCEMENT, *AUTOINSTRUCTIONAL AIDS, RESEARCH, COMMUNICATIONS SKILLS, LANGUAGE RESEARCH, *PERCEPTION, AUDITORY PERCEPTION, *ORAL COMMUNICATION, LISTENING, SPEECH AUTOINSTRUCTION DEVICE (SAID), ANN ARBOR, MICHIGAN

TEN EXPERIMENTAL ANALYSES RELATED TO THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION WERE REPORTED. THE ANALYSES WERE CONCERNED WITH THE DIFFERENTIAL EFFECTS OF STIMULI AND REINFORCEMENTS ON SPEECH PRODUCTION AND THE DISCRIMINATORY PERCEPTION OF SPEECH. INITIAL SPECIFICATIONS WERE INCLUDED FOR A SPEECH AUTOINSTRUCTION DEVICE TO BE USED FOR TEACHING PROSODIC FEATURES OF SPEECH. (RELATED REPORTS ARE ED 003 883-ED 003 887.) (WN)

ED003883

77
No. 3

04411-3-F

THE UNIVERSITY OF MICHIGAN

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
DEPARTMENT OF PSYCHOLOGY
BEHAVIOR ANALYSIS LABORATORY

Progress Report No. 3
February 1, 1962, to April 1, 1963

Experimental Analysis of the Control of Speech Production and Perception: III

Under contract with:

Language Development Section
U. S. Office of Education
Contract No. SAE-9265
Washington, D. C

Administered through:

April 1963

OFFICE OF RESEARCH ADMINISTRATION · ANN ARBOR

THE UNIVERSITY OF MICHIGAN
COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Department of Psychology
Behavior Analysis Laboratory

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the
person or organization originating it. Points of view or opinio
stated do not necessarily represent official Office of Educati
position or policy.

Progress Report No. 3

February 1, 1962, to April 1, 1963

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION: III

ORA Project 04411

under contract with:

LANGUAGE DEVELOPMENT SECTION
U. S. OFFICE OF EDUCATION
CONTRACT NO. SAE-9265
WASHINGTON, D.C.

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

April 1963

CONTENTS

DIFFERENTIAL REINFORCEMENT OF VOCAL DURATION

METRIC PROPERTIES OF MULTIDIMENSIONAL STIMULUS GENERALIZATION

THE EFFECTS OF RESPONSE-DEPENDENT AND INDEPENDENT REINFORCEMENT IN EXTENDING STIMULUS CONTROL

SOME DISCRIMINATIVE PROPERTIES OF SYNTACTIC STRUCTURES

RATIO SCALES, CATEGORY SCALES, AND VARIABILITY IN THE PRODUCTION OF LOUDNESS AND SOFTNESS

THE AUTOPHONIC SCALE OF VOICE LEVEL FOR CONGENITALLY DEAF SUBJECTS

A NEW DETERMINATION OF THE PURE TONE EQUAL-LOUDNESS CONTOURS FOR EARPHONE LISTENING

AN EFFECT OF NATIVE LANGUAGE AND MUSICAL TRAINING ON VOCAL PITCH MATCHING

JUDGMENT OF PORTRAYED EMOTION FROM NORMAL ENGLISH, FILTERED ENGLISH, AND JAPANESE SPEECH

RECOGNITION OF PORTRAYED EMOTION IN A FOREIGN LANGUAGE

INITIAL SPECIFICATIONS FOR A SPEECH AUTO-INSTRUCTIONAL DEVICE

APPENDIX

- STAFF
- ARTICLES PUBLISHED
- SPECIAL ACTIVITIES

DIFFERENTIAL REINFORCEMENT OF VOCAL DURATION

Harlan Lane
The University of Michigan

Differential Reinforcement of Vocal Duration[1]

Harlan Lane

Behavior Analysis Laboratory

University of Michigan

An excellent description of the procedure for differential reinforcement of responding has been given by Keller and Schoenfeld (1950):

"We select one (or more) of the 'natural' variations of a well-conditioned response and give it exclusive reinforcement; the remaining variations are subjected to extinction. If we pick out, in advance, a variation that has been of fairly frequent occurrence, and if we apply this selective reinforcement rigorously, we can soon produce an increase in the frequency of the response that possesses the property...we have chosen. At the same time we decrease the frequency of those responses that do not meet our specifications."

In the experiments to be reported, the well-conditioned response that was selected was the emission of the vowel /u/ by human adults. The natural variants, given exclusive reinforcement, were durations of the vocal response greater than a criterion duration. The value of the criterion was an experimental variable; in several studies we picked out variants that were fairly frequent, in others, some that were not as frequent. Selective reinforcement was applied rigorously to every variant above criterion. There was an increase in the frequency of responses whose duration exceeded criterion and a decrease in the frequency of those variants that went unreinforced.

Method

Measurement of duration

Techniques of measuring the parameters of the vocal operant have been discussed previously (Lane & Shinkman, 1963); the method employed to measure response duration will be summarized here. The vocal response was transduced by a dynamic microphone (Rye TM 1), amplified, and processed by an average speech power circuit (Peterson & McKinney, 1961). The output waveform was applied to a dc amplifier (Krohn-Hite) and then to the time-interval section of a frequency counter (Hewlett-Packard 523D). The duration was read in msec by the counter and recorded by a parallel printer and a card punch (IBM).

Subjects and procedure

Twenty male and eight female subjects (Ss) served individually in sessions lasting approximately two hours. The S was seated in a sound-treated booth in front of a panel containing a light and coin dispenser. These instructions were read to S:

"In this experiment you can earn money simply by saying /u/ when the light in front of you flashes. You will know that you have earned money when a penny falls into the tray; at the end of the experiment you may take home all the pennies you have earned. This is your only pay.

Occasionally the light will not flash for a period of time. When the light is off, please do not respond."

An electronic timer flashed the stimulus light at 5-sec intervals. After each of the first 20 responses emitted by S, the experimenter (E) presented reinforcement and entered the response duration on a desk calculator. Then, E reset the timer and computed the reinforcement criterion for the next phase of the experiment. In the second and later phases, only those responses whose duration exceeded the criterion were reinforced. Each phase was continued until 10 reinforced responses occurred in succession; then the timer was reset, a new criterion computed, and the next phase initiated.

Criteria for differential reinforcement

The duration criterion for reinforcement varied according to the schedule shown in Table 1. In Experiment I, the criterion for each phase equalled the mean duration of the 10 responses that terminated the preceding phase, plus their average deviation.

Following this experiment, the research detoured to obtain a preliminary estimate of the Weber fraction for produced vocal duration. Five Ss were instructed to produce 90 pairs of the vowel /u/ such that the responses in each pair had equal duration. Thirty pairs were obtained from each S at "short" durations, and a like number at "moderate" and "long" durations. When the absolute value of the difference in duration between the responses in each pair was plotted as a function of the shorter response, the 450 determinations were well fit by a straight line with a slope of 0.1 (determined by the method of least squares).[2]

In Experiments II through VI the criterion for reinforcement in each phase equalled the mean duration of the 10 terminal responses in the preceding phase plus or minus some per cent of that mean. The step size or increment in the criterion ranged from half the value of the Weber fraction (5 per cent) to twice that value (20 per cent).

Probing for a tact of the reinforcement contingency

At the end of each session in Experiments II - VI, S was given a three-item questionaire. The response to each item was completed before the next was presented. The questions were: (1) What do you think was the point of this experiment? (2) Did you know what it was about your responding that earned pennies for you? (3) Did you know that you received pennies depending on the length or duration of your response? If the words "duration" or "length" never appeared in _S_'s response and if the word "no" was part of the response to questions (2) and (3), the questionaire was scored "negative".

Results and Discussion

Experiment I: Shaping longer durations; step size constant.

The relation between the criterion for differential reinforcement and the mean response duration in each phase is shown in Fig. 1 for a representative _S_. The terminal 10 reinforced responses, which ended each phase (filled triangles), the non-terminal reinforced responses (open triangles) and the unreinforced responses (inverted triangles) are plotted separately.

It is apparent that the duration of vocal responding by S5 was under the control of the reinforcement contingencies. As the criterion was increased, in steps of one average deviation, from 270 to 570 msec, the mean duration of vocal responding was shaped from 258 to 610 msec. In each phase, the mean duration of unreinforced, non-terminal and terminal reinforced responses increased in that order. This finding largely reflects the increasing duration of vocal responding within the phase, which is shown for a typical phase (the last) in the insert. Note that the last 10 responses in this phase had a mean duration 40 msec greater than the criterion. This "overshoot" is characteristic of the other phases as well. If the mean response durations precisely equalled the criterion, they would appear in the graphs along a line with unit slope, which passes through the intersections of corresponding points on the x- and y- axes. In general, the non-terminal reinforced responses fall near this line, that is, at criterion values, while reinforced responses lie above, and unreinforced responses lie below the line.

Frequency distributions of response durations in each phase are shown for S5 in Fig. 2. Under continuous reinforcement, response durations are short and the variance is relatively small. The criterion for each phase and for the one that follows are indicated by the symbols $\underline{C}$ and $\underline{C}_1$, respectively. With differential reinforcement, the variance increases and, in the later phases, the frequency distributions become bimodal. For most of the $\underline{S}$s, the second peak lies at durations greater than the criterion and reflects the "overshoot" occurring toward the end of each phase, which was described above.

Figure 2 also shows that S5 reached the successive criteria rapidly; the number of unreinforced responses ranged from 5 in phase 2 to a maximum of 20 in the last phase. A useful statistic to consider in this regard is the difference between the criterion duration and the mean duration of the last phase, divided by the standard deviation of that mean: $(C-M)/\sigma$. This statistic, which we may conveniently call the shaping index, expresses the step size (C-M) in terms of the variability (σ) of current behavior. An infrequent variant of behavior has been selected for differential reinforcement if the shaping index is large, either because the step size is large, or the variability is small, or both. For the 16 $\underline{S}$s whose data are presented in this paper, the shaping index computed for each phase correlated .4 with the number of unreinforced responses in that phase (Spearman rho, $p < .01$). The shaping index for S5 is roughly constant at about 0.7; this follows from the fact that the criterion increment was one average deviation.

The relation between response duration and the reinforcement criterion is shown for a second S in Fig. 3. The shaping index was constant at about 0.7; the criterion ranged from 272 to 606 msec while the average response duration ranged from 256 to 736 msec. Overshoot of the criterion was particularly marked in the last phase, which is shown in greater detail in the insert. There were 72 unreinforced responses, all of which had relatively short durations, continuing a trend established earlier in the session. The mean of the non-terminal reinforced responses was 20 msec greater than criterion, as in earlier phases. The terminal responses in the last phase had a mean duration that exceeded criterion by 130 msec.

Experiment II: Shaping longer durations, step size increasing, then shorter durations, step size increasing.

Mean response duration as a function of criterion duration is shown for a representative S in Fig. 4; the corresponding frequency distributions of response durations appear in Fig. 5. The mean duration grows steadily during the first five phases and the variability increases proportionally ($\sigma/M = .05$). In phase 6, the criterion for reinforcement changed sign: any response duration less than 446 msec was followed by reinforcement. The relative variability doubled ($\sigma/M = .10$); the mean duration of the unreinforced responses climbed abruptly to 570 msec; while that of the terminal responses fell to 334 msec, an overshoot of 112 msec. The frequency distribution of response durations spread and became trimodal, with peaks corresponding, roughly, to the terminal, non-terminal reinforced and unreinforced responses, respectively. Order was restored rapidly in the next phase, however, where the criterion increment was small and of the same sign. The shaping index for this S varied from 1.1 in phase 2 to a maximum of 3.0 in the last phase. The latter value is large because the variability in the preceding phase was small and the step size was 20 per cent of the mean. There were nearly four times as many unreinforced responses in this phase as in any other. The questionaire, which probed for a tact of the reinforcement contingency, was scored "negative" for this S.

The data for S13 (Fig. 6) provide a similar picture of the shaping of vocal duration. The criterion ranged upward from 369 to 1353 msec and then downward to 433 msec; the mean duration of the terminal responses ranged from 348 to 1424 to 285 msec. The largest shaping index (2.8) occurred in phase 5 and there was a corresponding increase in the number of unreinforced responses (135, or 10 times more than in any other phase). When the sign of the criterion changed in phase 6, an abrupt increase in the duration of unreinforced responses occurred. This finding is obtained whenever the sign of the criterion is changed; it may be related to the observation, reported by Lane and Shinkman (1963), that several parameters of the vocal operant, including duration, increase in extinction after intermittent reinforcement. In all phases but the fifth, criterion was reached rapidly by S13, who described the reinforcement contingency on the first sheet of the questionaire.

The findings obtained with three Ss who did not reach criterion in all phases are examined next (Figs. 7 and 8). The data for S15 are anomalous in several respects. The mean duration of the non-terminal reinforced responses lies above that of the terminal responses in phases 2 and 3. The shaping index was quite large (3.2) in phase 4, but criterion was reached within 50 responses. The 20 per cent increment in criterion in phase 5 was not a large step, in view of the variability in the preceding phase, but S15 failed to receive a single reinforcement in this phase.

The failure to shape the behavior of S14 (Fig. 8) may be largely attributed to the size of the shaping index in the uncompleted phase: 3.8. It is interesting to note that the mean duration of the 290 unreinforced responses in the last phase fell below that mean for the preceding phase. The size of the shaping index similarly accounts for the breakdown of reinforcement control with S20. The step size in phase 4 was moderately large (15 per cent) and there was little variability in the preceding phase; hence, the shaping index took on the exceedingly large value, 8.4; 208 unreinforced responses occurred before the session was terminated by E. The questionaires for Ss 14, 15, and 20 were all scored "negative."

Experiment III: Shaping longer durations, step size increasing, then shorter durations, step size decreasing.

The first half of this procedure (phases 1-5) replicates that for Experiment II and the behavior of S16 (Figs. 9, 10) is similar in these phases to that of S12 and S13, presented earlier. The frequency distributions of response durations have greater variance for S16 than for earlier Ss. In phase 6, when the step size was both large (20 per cent) and of opposite sign, the frequency distribution separated into two clusters; one reflects the marked increase in the duration of unreinforced responses, the other the very short durations of the terminal responses, whose mean, 279 msec, overshot the criterion by 487 msec. The shaping index of 3.3 and the number of unreinforced responses were both larger for phase 6 than for any other phase. The terminal response duration in phases 7 - 9 was less than the mean duration observed under CRF; it reached as low as 59 msec in the last phase, 157 msec less than the baseline determination.

The findings for S19 (Fig. 11) are similar, although the range of response durations is much smaller: 165 to 532 to 94 msec. The shaping index was approximately 1.2 except in phase 5 where, coupled with a 20 per cent step size, the index rose to 4.1; the number of unreinforced responses rose from an average of 25 to a high of 185 in this phase. The mean duration of the terminal responses in phase 9 was 94 msec or 71 msec below the baseline established under CRF. Both S16 and S19 described the reinforcement contingency on the questionaire.

Experiment IV: Shaping longer, then shorter, durations; step size small and constant.

When the shaping index remains small, as in the present procedure, variability remains small, shaping proceeds gradually and the number of steps required to change response parameters over a given range is increased. Subject 21, for example, shows a steady increase in mean duration (Fig. 12) during the first five phases and then a rapid decline when shorter durations were reinforced. The step size was five per cent. The shaping index ranged from .4 in the first and last phases to approximately 1.0 in the others. Variability remained relatively small, $\sigma/M = .07$, with the notable exception of phase 6, where the sign of the criterion changed. Subject 21 responded "yes" to items (2) and (3) on the questionaire.

The findings for S23 (Fig. 13) provide a second example of the result of maintaining a small shaping index. In this case, the index ranged from .3 to a maximum of 1.0 in phase 6, where the criterion changed sign. Variability remained small, $\sigma/M = .06$, and the range of durations for the reinforced responses was limited: 246 to 519 to 300 msec. Shaping took place quite rapidly. In the final phase, the reinforcement criterion was reached immediately, with no unreinforced or non-terminal reinforced responses occurring. The questionaire for this S was scored negative.

Experiment V: Shaping longer durations; step size large and constant.

When a step size as large as 20 per cent was employed at the outset of the shaping procedure, Ss 24 and 25, whose data are presented in Fig. 14, failed to reach criterion after the first phase of differential reinforcement. The baseline duration for S24 was 174 msec; he reached criterion in the second phase, when the shaping index was 2.6. In the third phase, however, the shaping index rose to 5.0 and no reinforced responses occurred. The findings for S25 are similar, although all the values are larger since his baseline duration was 327 msec. The shaping index in the third phase was 3.8. A total of 273 responses were emitted in this phase, including 24 reinforced responses. It is interesting to note that the last ten of these had a mean duration 26 msec greater than criterion. However, numerous unreinforced responses were interspersed and the criterion of 10 successive responses with duration greater than 520 msec was never reached. The questionaire for S24 was scored negative whereas that for S25 was not.

Experiment VI: Shaping shorter durations; step size increasing.

The attempt to shape downward from the operant baseline was successful only over a limited range. The findings for three Ss are shown in Fig. 15. In phases 2 and 3 the shaping indices for all three Ss ranged from .5 to 1.5 and differential reinforcement was effective. At the end of phase 3, the mean duration and variability were quite small for all Ss. This finding is shown graphically for S27 in Fig. 16. As a result, the shaping index for the next phase was relatively large, approximately 2.5 for the three Ss. No reinforced responses were emitted by Ss 26 and 28 in phase 4. Three responses by S27 were reinforced but this did not lead to control over the duration of vocal responding. The questionaires for all three Ss were scored negative.

Summary

The effects of differential reinforcement of vocal duration were examined in a series of experiments in which a total of 28 subjects emitted a single vowel each time a light was flashed. After establishing a baseline duration under continuous reinforcement, (in which pennies were dispensed to S), the experimenter selectively reinforced those responses with durations greater than criterion. When ten successive responses were reinforced, a new criterion was determined and the experiment continued. The criterion for reinforcement was varied in the six experiments; it ranged from 80 to 120 per cent of the mean duration of precurrent responding, according to one of six schedules.

Differential reinforcement effected a large and systematic change in the duration of vocal responding as long as the responses selected for reinforcement had a sufficiently high probability of occurrence. This requirement was formulated as the difference between the criterion duration and the mean duration of current responding, divided by the standard deviation of that mean. This statistic, named the shaping index, was correlated with the number of responses that were emitted before criterion was reached, and it was found to be large whenever the shaping process failed.

Many of the subjects, whose vocal durations were changed radically during the experiment, failed to tact the reinforcement contingency despite extensive probing by a questionaire, administered at the end of each session.

References

Keller, F. S. & Schoenfeld, W. N. Principles of psychology. New York: Appleton-Century-Crofts, 1950.

Lane, H. L. & Shinkman, P. G. Methods and findings in an analysis of a vocal operant. J. exp. Anal. Behav., 1963, 6, 179-188.

Peterson, G. E. & McKinney, N. P. The measurement of speech power. Phonetica, 1961, 7, 65-84.

Footnotes

1. This research was performed pursuant to a contract with the Language Development Section, United States Office of Education.

2. The value of the Weber fraction, 10 per cent, turned out to coincide with the average coefficient of variation (σ/M) for the duration of responses emitted under continuous reinforcement.

Table 1

SUCCESSIVE CRITERIA FOR DIFFERENTIAL REINFORCEMENT OF VOCAL DURATION

Expt. No.	Subjects	Step Size	Phase Number								
			1	2	3	4	5	6	7	8	9
I	1-10	+	CRF	M_1+AD_1	M_2+* AD_2	M_3+AD_3	M_4+AD_4	M_5+AD_5	M_6+AD_6		
II	11-15 & 20	+,+	CRF	$M_1+.05M_1$	$M_2+.10M_2$	$M_3+.15M_3$	$M_4+.20M_4$	$M_5-.05M_5$	$M_6-.10M_6$	$M_7-.15M_7$	$M_8-.20M_8$
III	16-19	+,-	CRF	$M_1+.05M_1$	$M_2+.10M_2$	$M_3+.15M_3$	$M_4+.20M_4$	$M_5-.20M_5$	$M_6-.15M_6$	$M_7-.10M_7$	$M_8-.05M_8$
IV	21-23	Const.	CRF	$M_1+.05M_1$	$M_2+.05M_2$	$M_3+.05M_3$	$M_4+.05M_4$	$M_5-.05M_5$	$M_6-.05M_6$	$M_7-.05M_7$		
V	24-25	Const.	CRF	$M_1+.20M_1$	$M_2+.20M_2$	$M_3+.20M_3$					
VI	26-28	+	CRF	$M_1-.05M_1$	$M_2-.10M_2$	$M_3-.15M_3$					

* Mean duration of 10 terminal responses in phase 2 plus their average deviation.

Figure Captions

Fig. 1. Shaping longer vocal duration; step size constant (Experiment I, S5). The relation between the criterion response duration for differential reinforcement and the mean response duration. Unreinforced responses (inverted triangles), non-terminal reinforced responses (open triangles) and the 10 terminal reinforced responses (filled triangles) are plotted separately for each phase of the experiment. The baseline duration under continuous reinforcement is shown by the symbol at the lower left. The vertical lines represent one standard deviation. The insert shows the mean duration of successive blocks of 10 responses in the last phase.

Fig. 2. Frequency distributions of response durations in each phase, for S5. The criterion duration for a phase is indicated by C, that for the next phase by C_1.

Fig. 3. The relation between the criterion and mean response duration, for S8 (Experiment I). The insert shows the mean duration of successive blocks of 20 responses in the last phase.

Fig. 4. Shaping longer durations, step size increasing, then shorter durations, step size increasing (Experiment II, S12). The relation between criterion and response duration. In phase 6, the sign of the criterion was changed (-), and all responses with duration less than the criterion were reinforced.

Fig. 5. Frequency distributions of response durations for S12. The negative symbols denote reinforcement of responses with duration less than C.

Figure Captions (Cont.)

Fig. 6. The relation between criterion and response duration for S13 (Experiment II).

Fig. 7. The relation between criterion and response duration for S15 (Experiment II). In phase 5, S15 failed to reach criterion.

Fig. 8. The relation between criterion and response duration for _S_s 14 and 20 (Experiment II). S14 failed to reach criterion in phase 3, S20 in phase 4. The symbols at the break in the abscissa indicate baseline durations for the two _S_s.

Fig. 9. Shaping longer durations, step size increasing, then shorter durations, step size decreasing (Experiment III, S16). The relation between criterion and response duration.

Fig. 10. Frequency distributions of response durations for _S_16.

Fig. 11. The relation between criterion and response duration for S19 (Experiment III).

Fig. 12. Shaping longer, then shorter durations; step size small and constant (Experiment IV, S21). The relation between criterion and response duration.

Fig. 13. The relation between criterion and response duration for S23 (Experiment IV).

Fig. 14. Shaping longer durations; step size large and constant (Experiment V, _S_s 24, 25). The relation between criterion and response duration. S24 emitted no reinforced responses in the third phase. S25 emitted 24 reinforced responses, including 10 that exceed criterion (arrow) but unreinforced responses were interspersed and criterion was not reached.

Figure Captions (Cont.)

Fig. 15. Shaping shorter durations; step size increasing (Experiment VI, Ss 26, 27, 28). The relation between criterion and response duration. None of the Ss reached criterion in the fourth phase.

Fig. 16. Frequency distributions of response durations for S27 (Experiment VI).

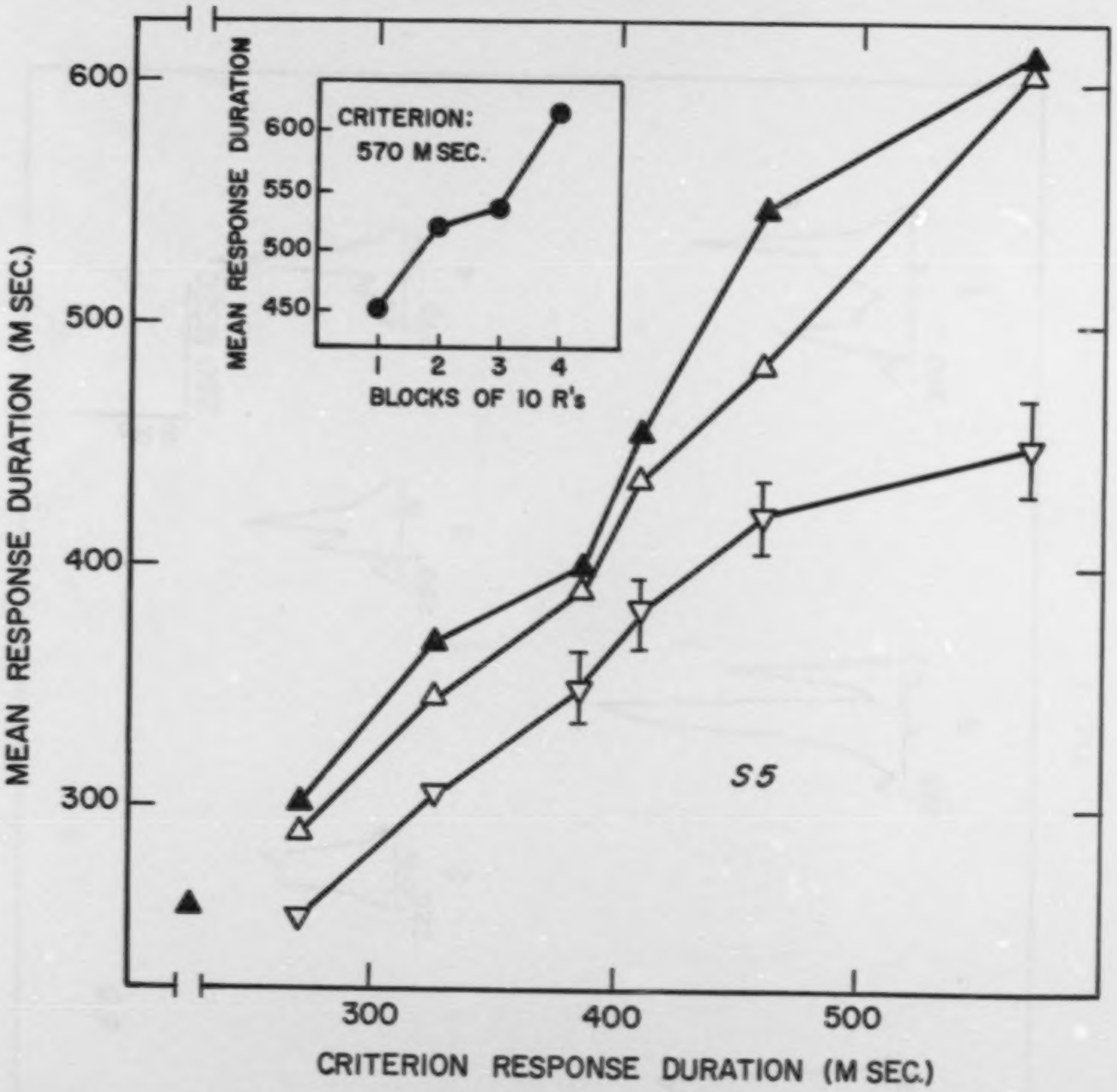

MEAN RESPONSE DURATION (M SEC.)
600
500
400
300
CRITERION RESPONSE DURATION (M SEC.)
300
400
500
S5
CRITERION:
570 M SEC.
MEAN RESPONSE DURATION
600
550
500
450
1
2
3
4
BLOCKS OF 10 R's

Fig. 1

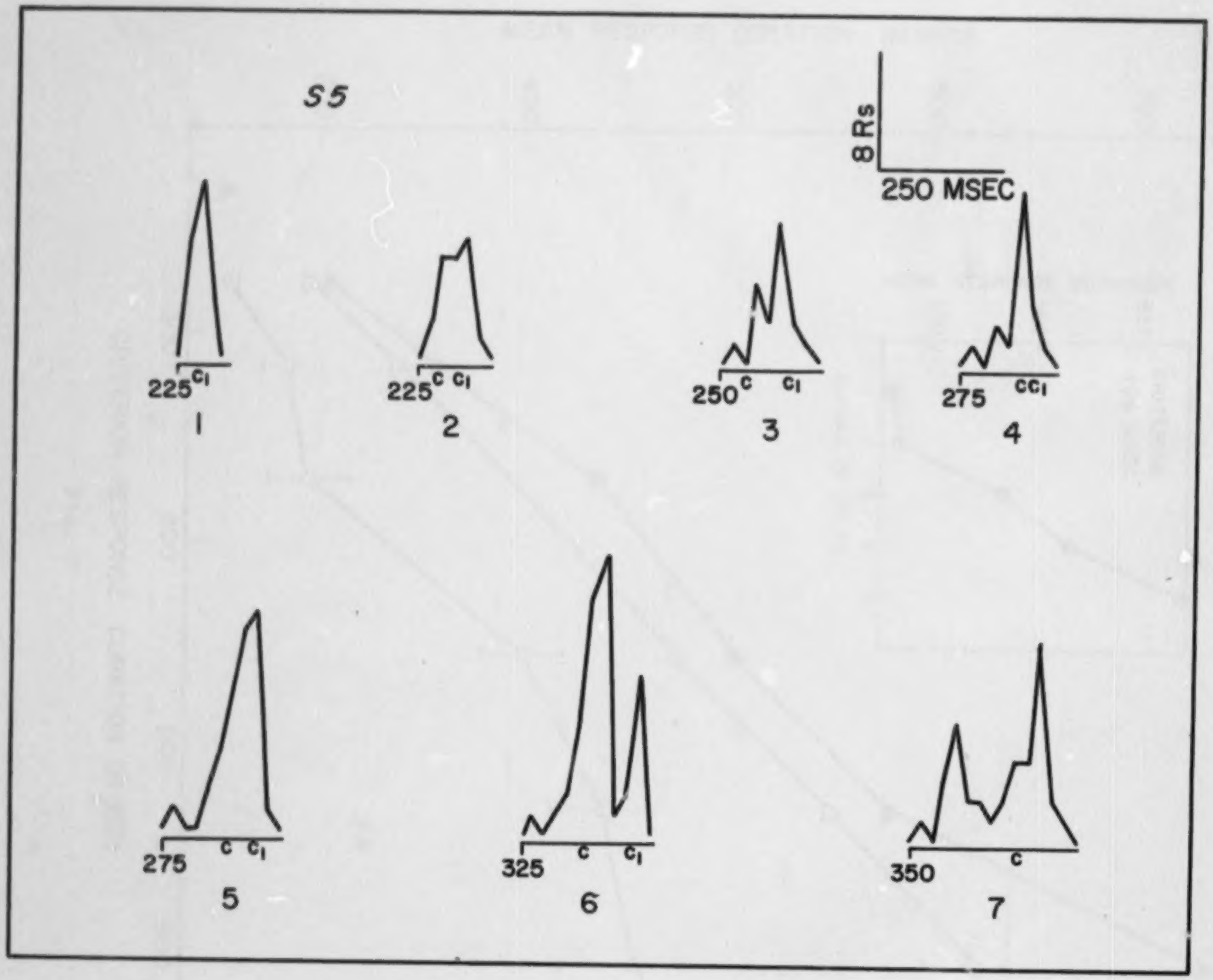
S5
8 Rs
250 MSEC
225
C1
1
225
C C1
2
250
C C1
3
275
CC1
4
275
C C1
5
325
C C1
6
350
C
7

Fig. 2

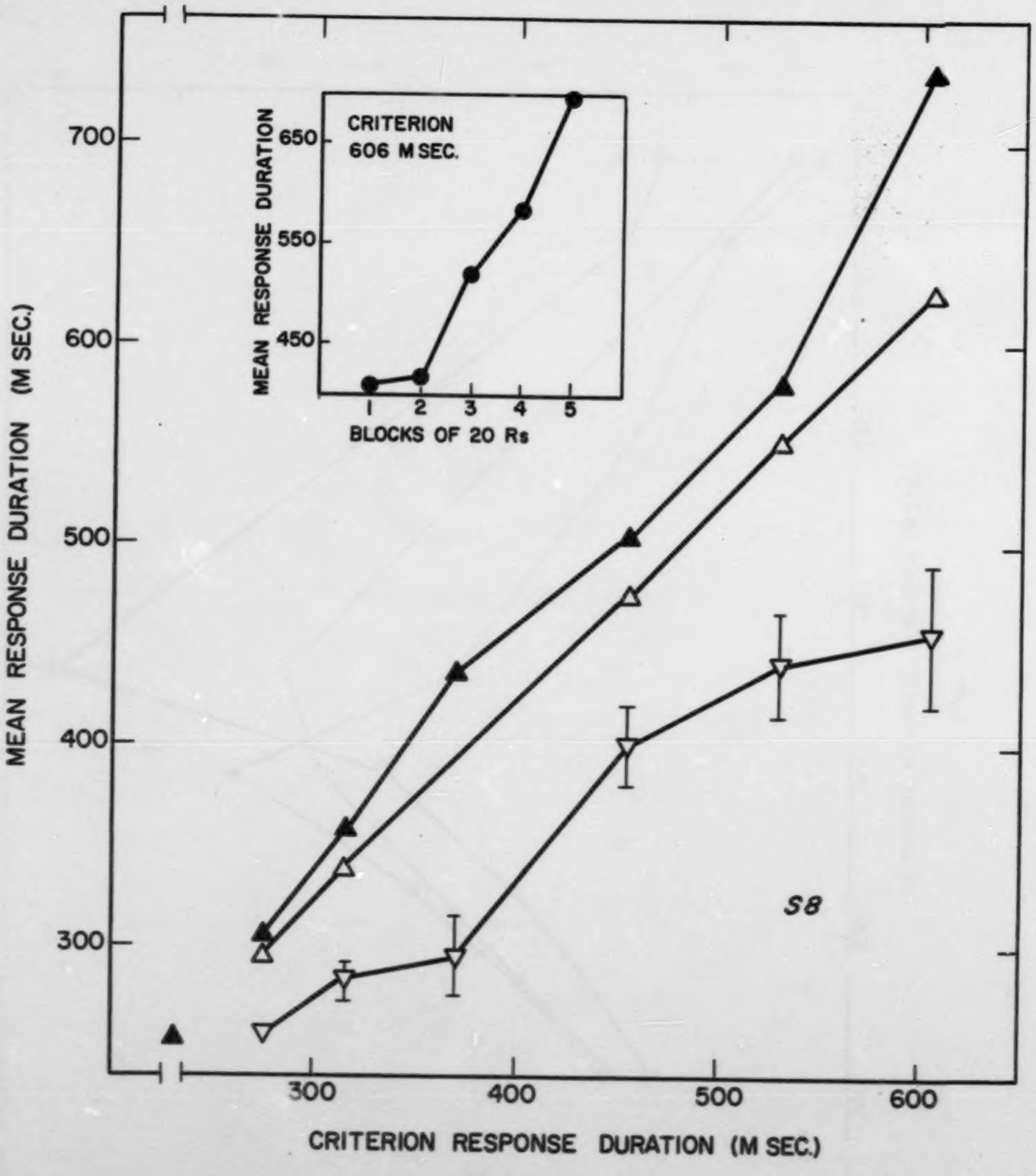

MEAN RESPONSE DURATION (M SEC.)
700
600
500
400
300
CRITERION RESPONSE DURATION (M SEC.)
300
400
500
600
S8
CRITERION
606 M SEC.
MEAN RESPONSE DURATION
650
550
450
1
2
3
4
5
BLOCKS OF 20 Rs

Fig. 3

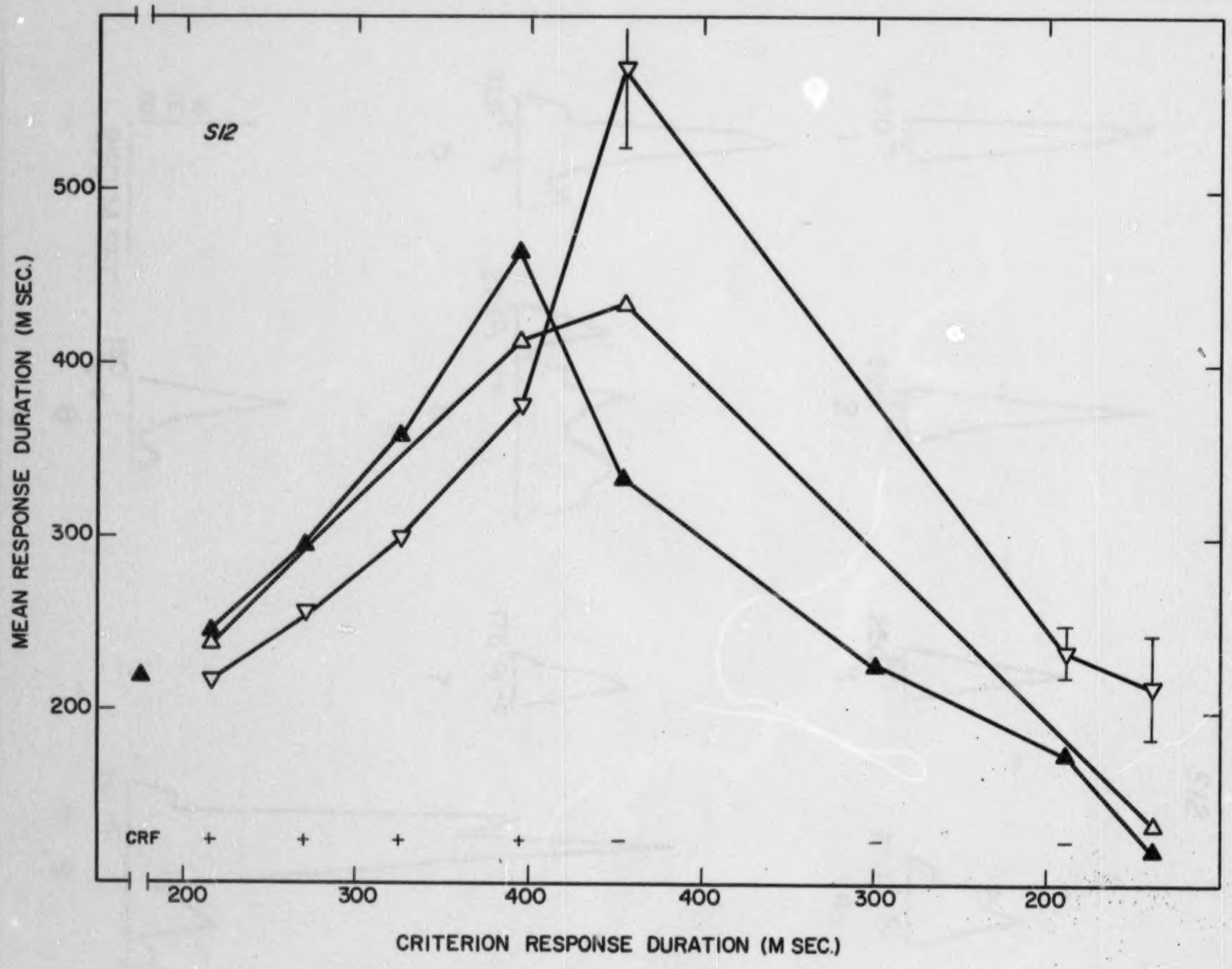

S12
MEAN RESPONSE DURATION (M SEC.)
500
400
300
200
CRF
+
+
+
+
−
−
−
200
300
400
400
300
200
CRITERION RESPONSE DURATION (M SEC.)

Fig. 4

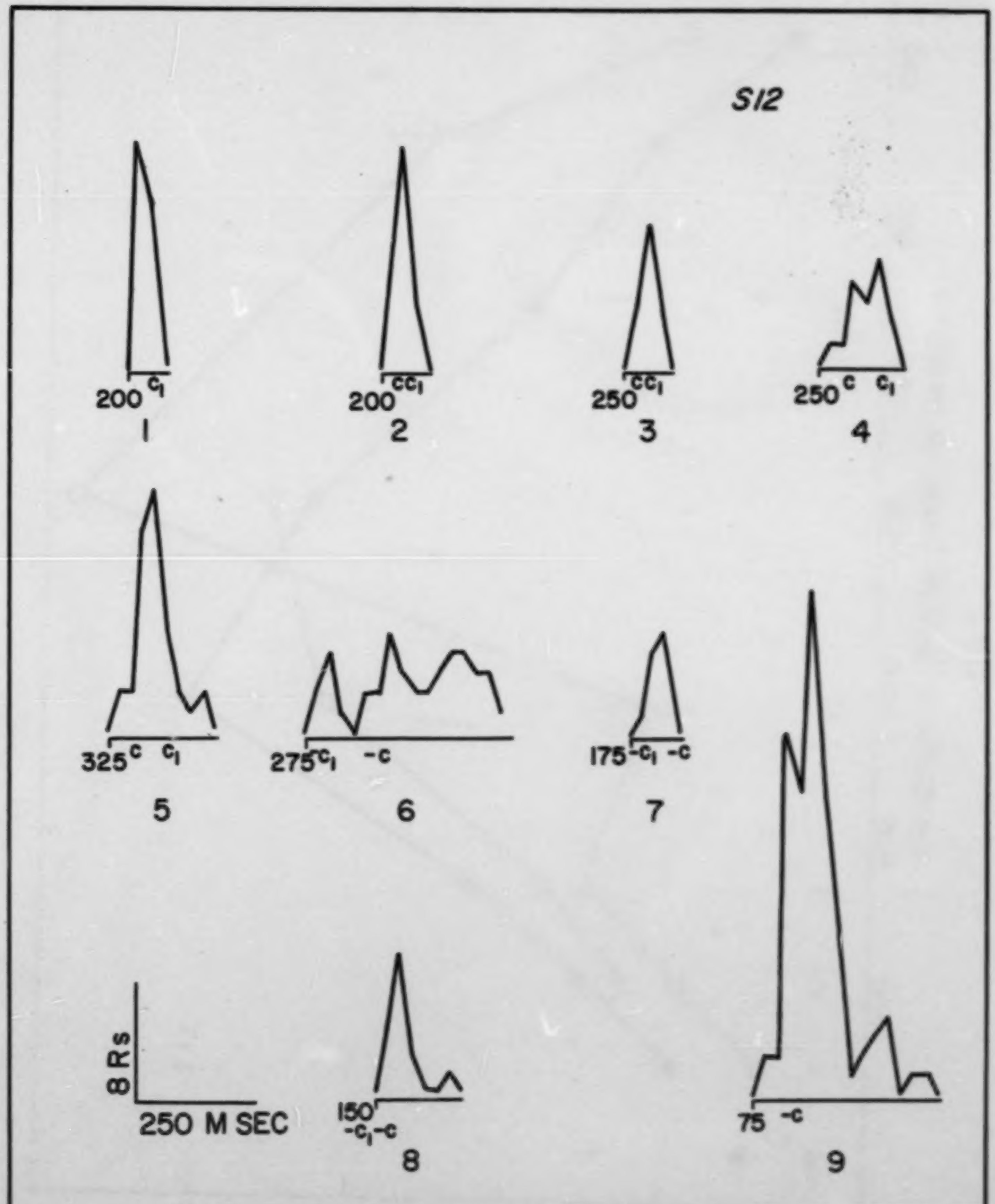

S12
200 c1
1
200 cc1
2
250 cc1
3
250 c c1
4
325 c c1
5
275 c1 -c
6
175 -c1 -c
7
8 Rs
250 M SEC
150 -c1-c
8
75 -c
9

Fig. 5

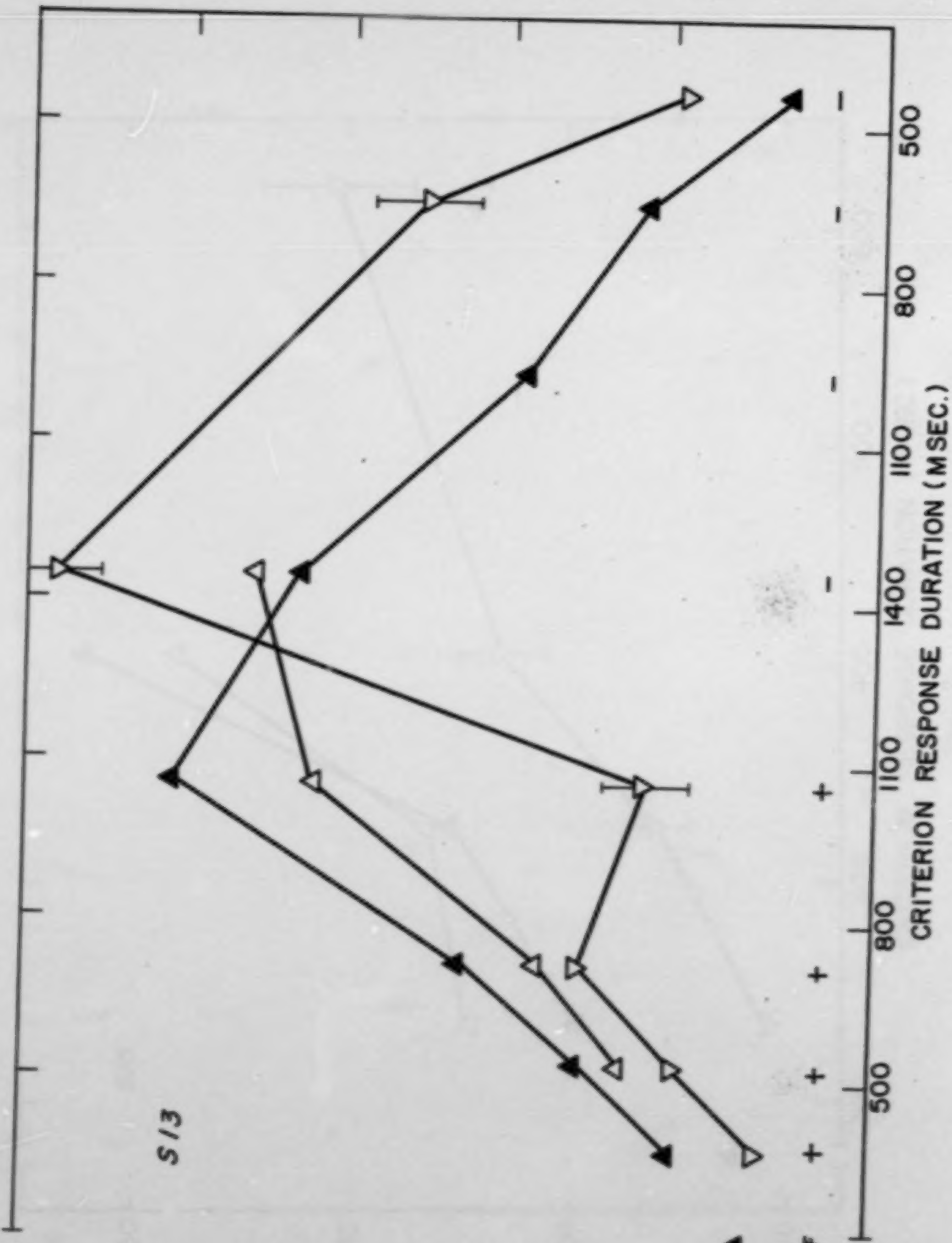

S/3
MEAN RESPONSE DURATION (MSEC.)
500
800
1100
1400
CRF
+
+
+
+
−
−
−
−
500
800
1100
1400
1100
800
500
CRITERION RESPONSE DURATION (MSEC.)

Fig. 6

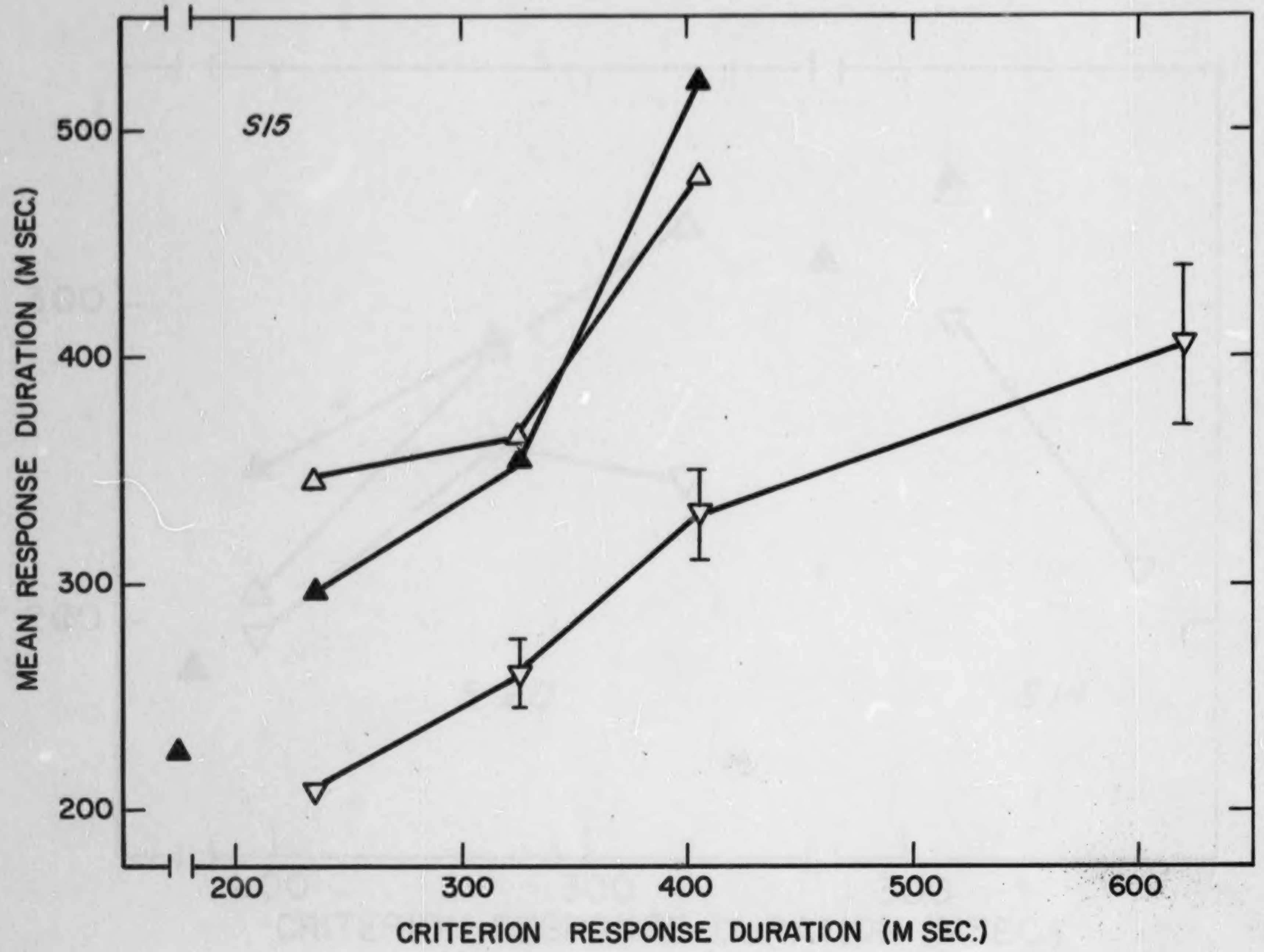

S/5
MEAN RESPONSE DURATION (M SEC.)
500
400
300
200
CRITERION RESPONSE DURATION (M SEC.)
200
300
400
500
600

Fig. 7

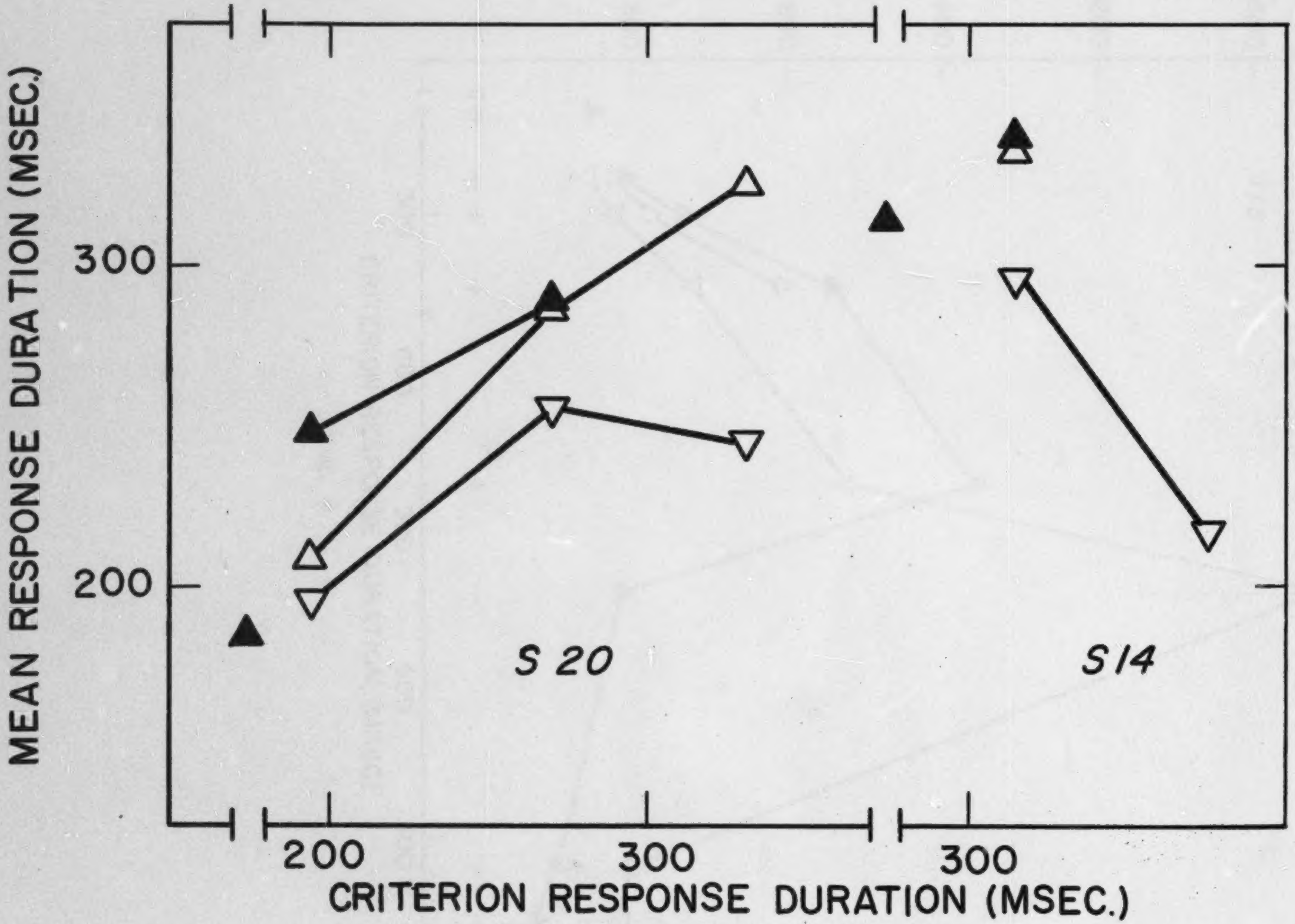

MEAN RESPONSE DURATION (MSEC.)
300
200
S 20
S 14
200
300
300
CRITERION RESPONSE DURATION (MSEC.)

Fig. 8

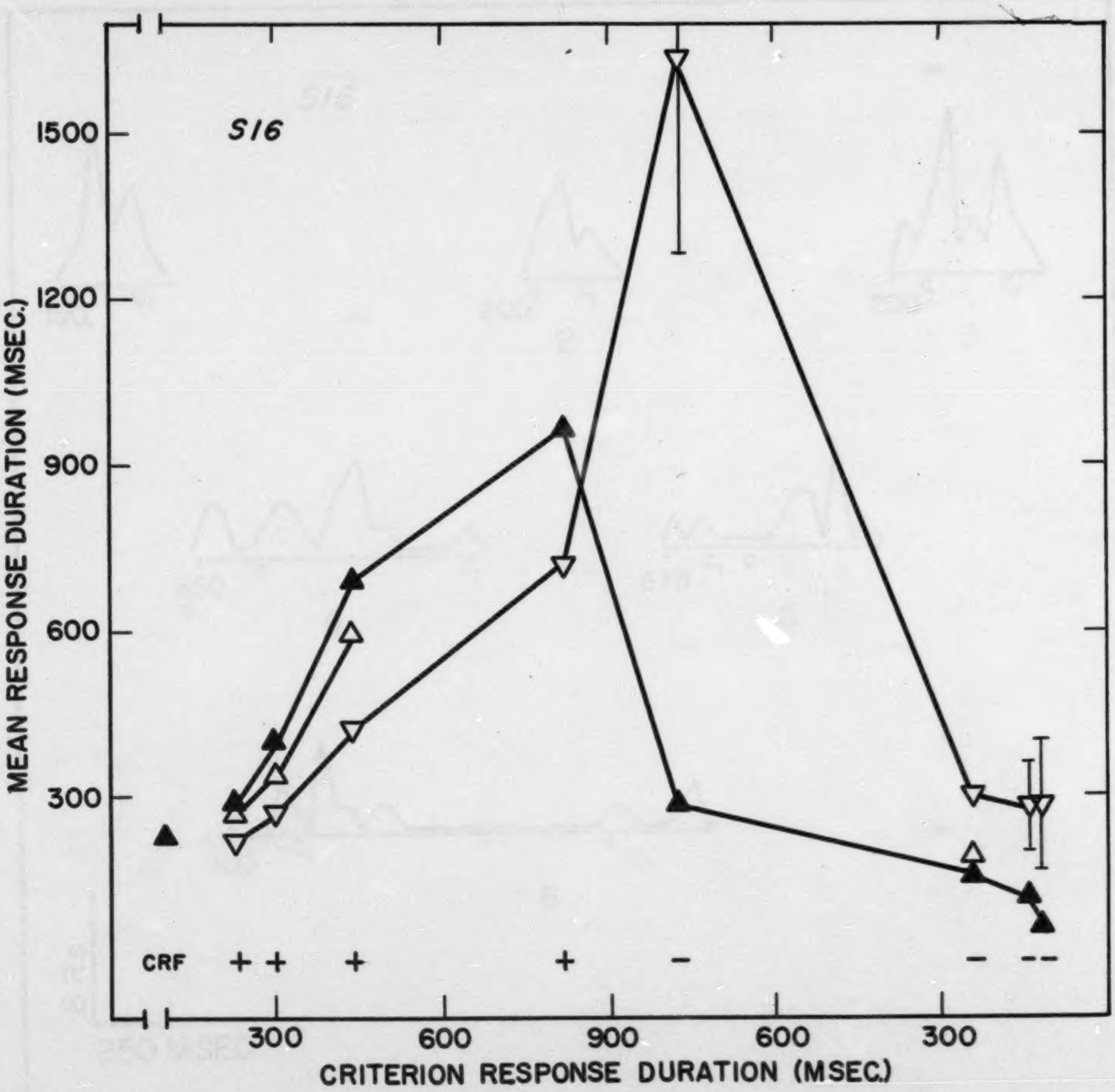

S16
1500
1200
900
600
300
MEAN RESPONSE DURATION (MSEC.)
CRF
+ +
+
+
–
–
– –
300
600
900
900
600
300
CRITERION RESPONSE DURATION (MSEC.)

Fig. 9

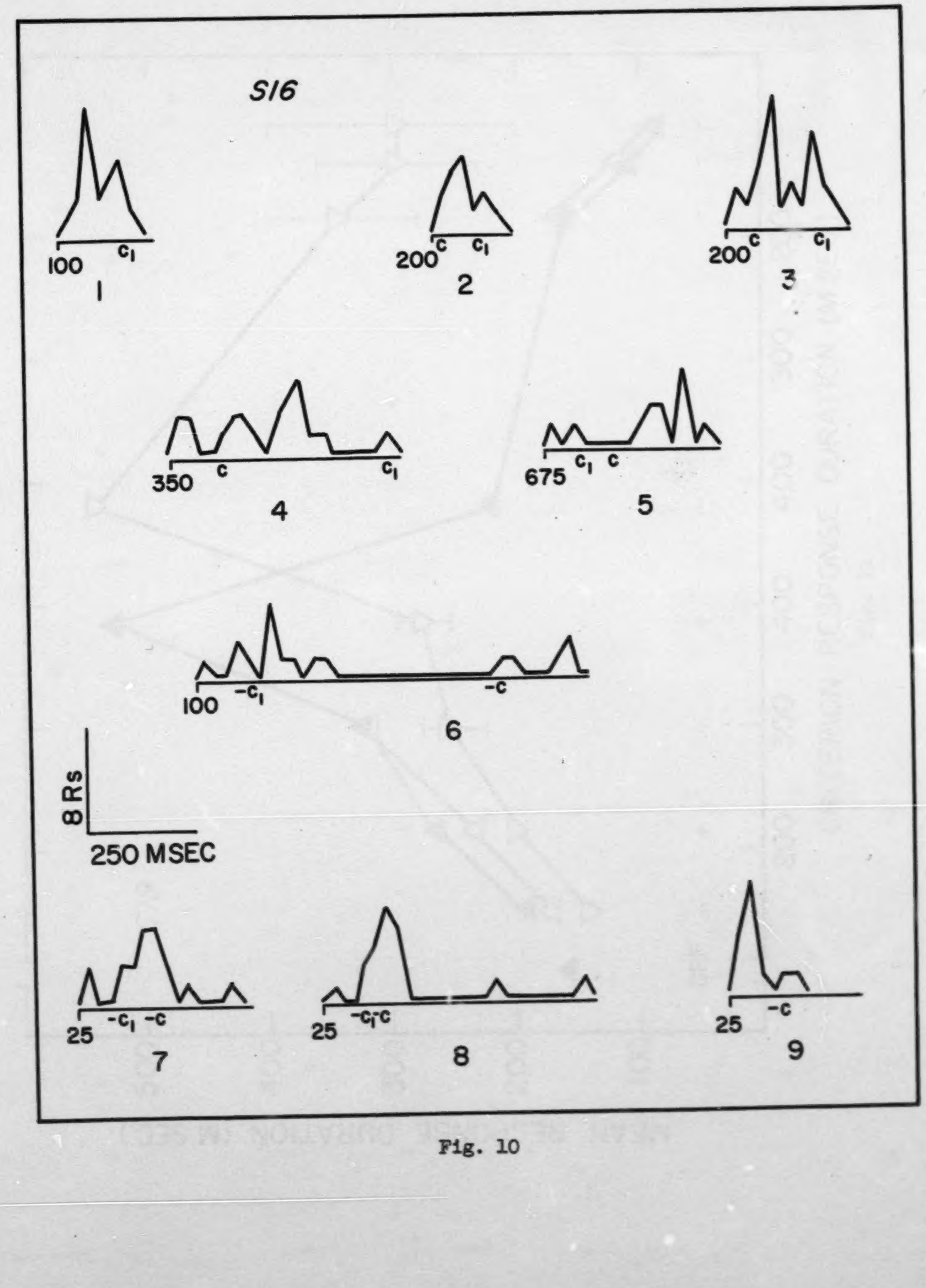

S16
100
c1
1
200
c
c1
2
200
c
c1
3
350
c
c1
4
675
c1
c
5
100
-c1
-c
6
8 Rs
250 MSEC
25
-c1
-c
7
25
-c1-c
8
25
-c
9

Fig. 10

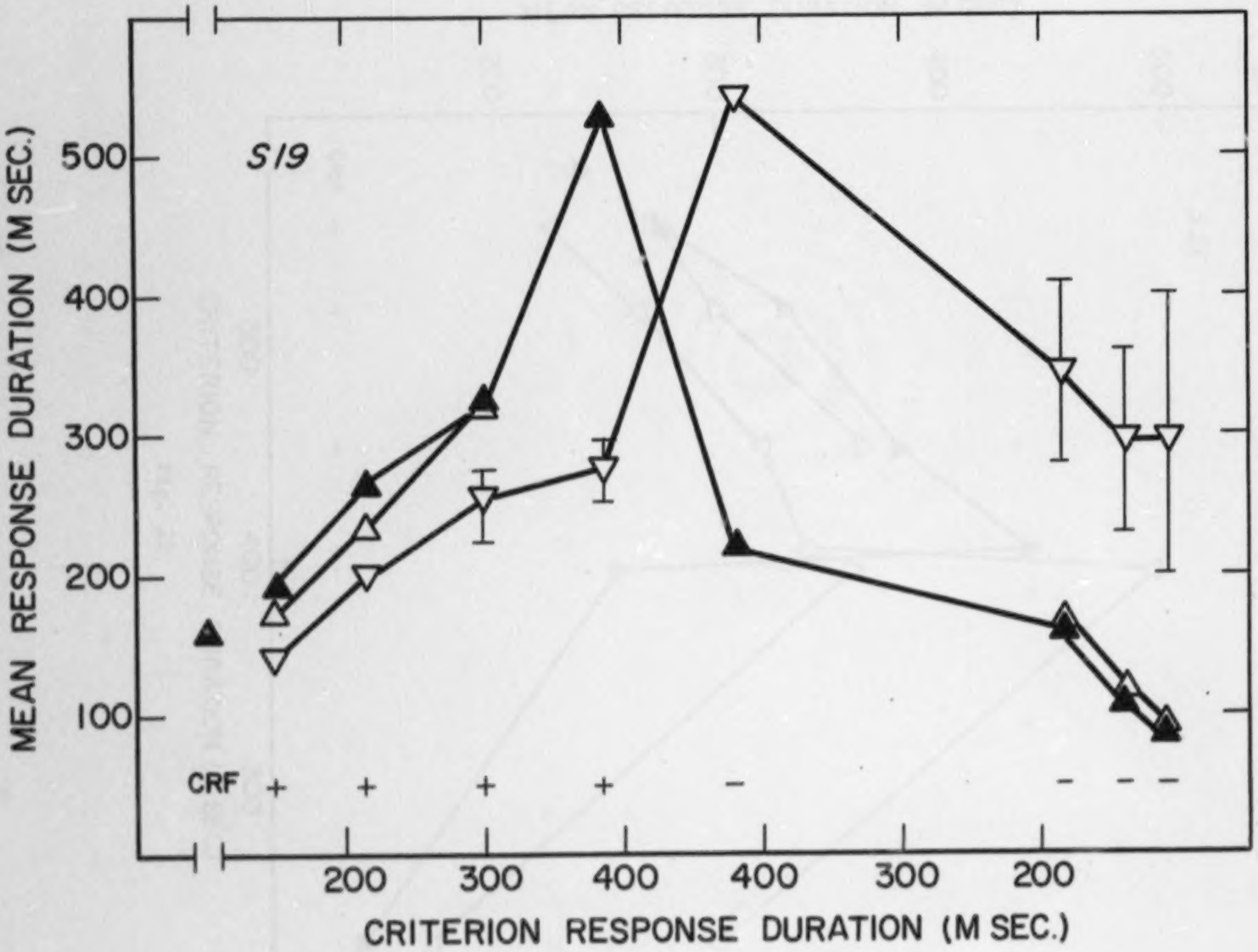

S19
MEAN RESPONSE DURATION (M SEC.)
500
400
300
200
100
CRF
+
+
+
+
−
−
−
−
200
300
400
400
300
200
CRITERION RESPONSE DURATION (M SEC.)

Fig. 11

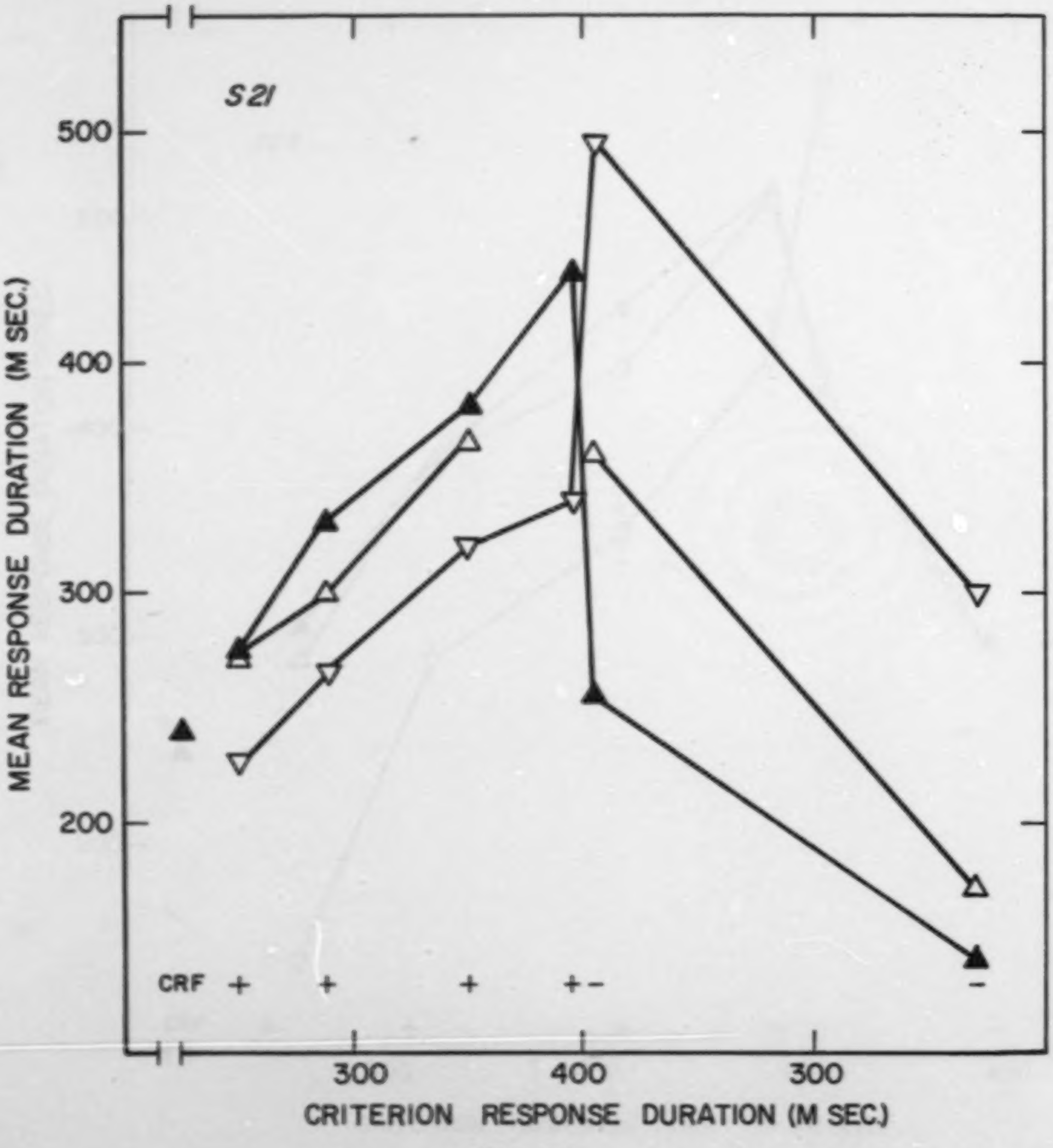

S21
500
400
300
200
MEAN RESPONSE DURATION (M SEC.)
CRF + + + + - -
300
400
300
CRITERION RESPONSE DURATION (M SEC.)

Fig. 12

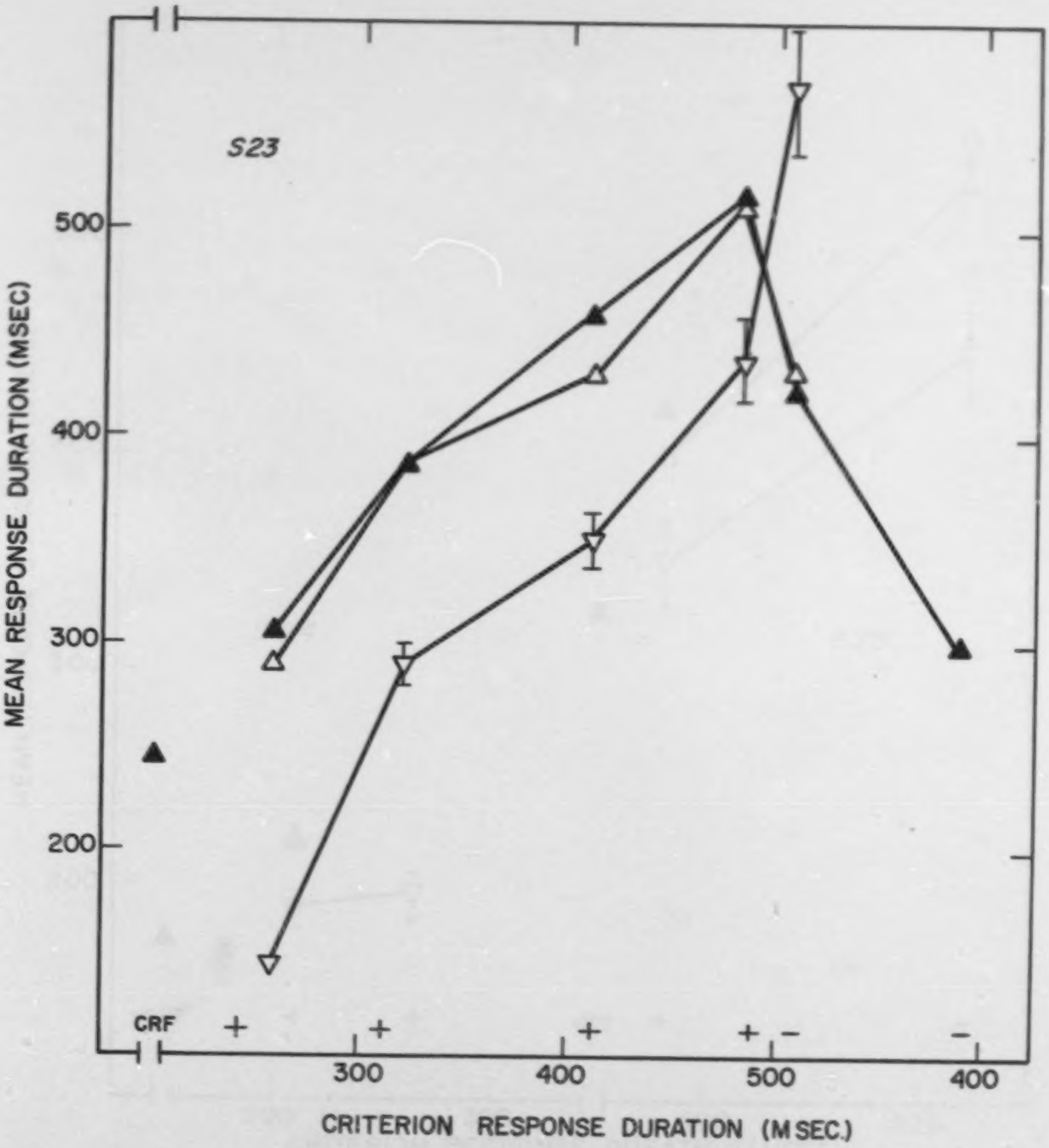

S23
MEAN RESPONSE DURATION (MSEC)
500
400
300
200
CRF
+
+
+
+
–
–
300
400
500
400
CRITERION RESPONSE DURATION (MSEC.)

Fig. 13

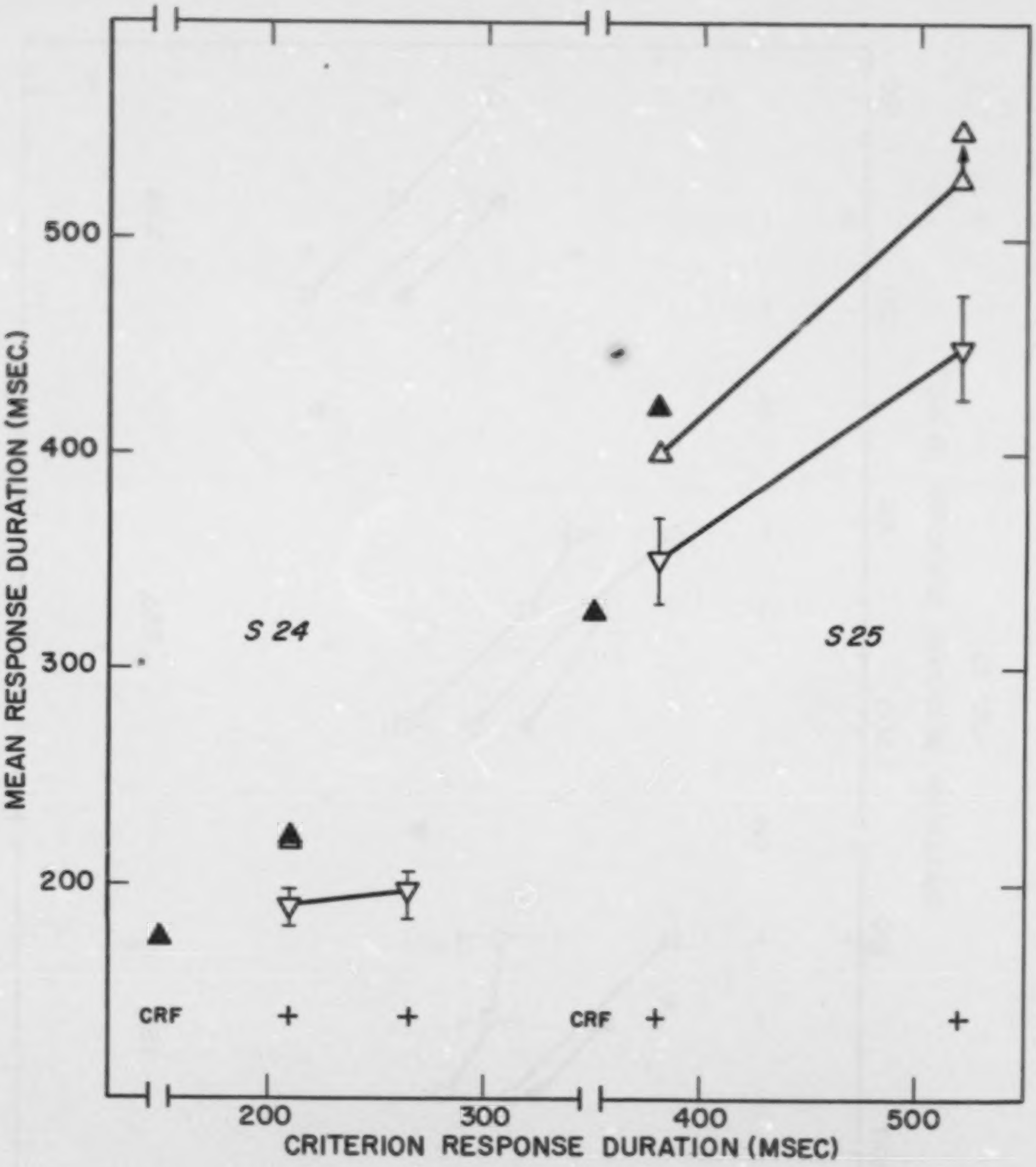

MEAN RESPONSE DURATION (MSEC.)
500
400
300
200
S 24
S 25
CRF
+
+
CRF
+
+
200
300
400
500
CRITERION RESPONSE DURATION (MSEC)

Fig. 14

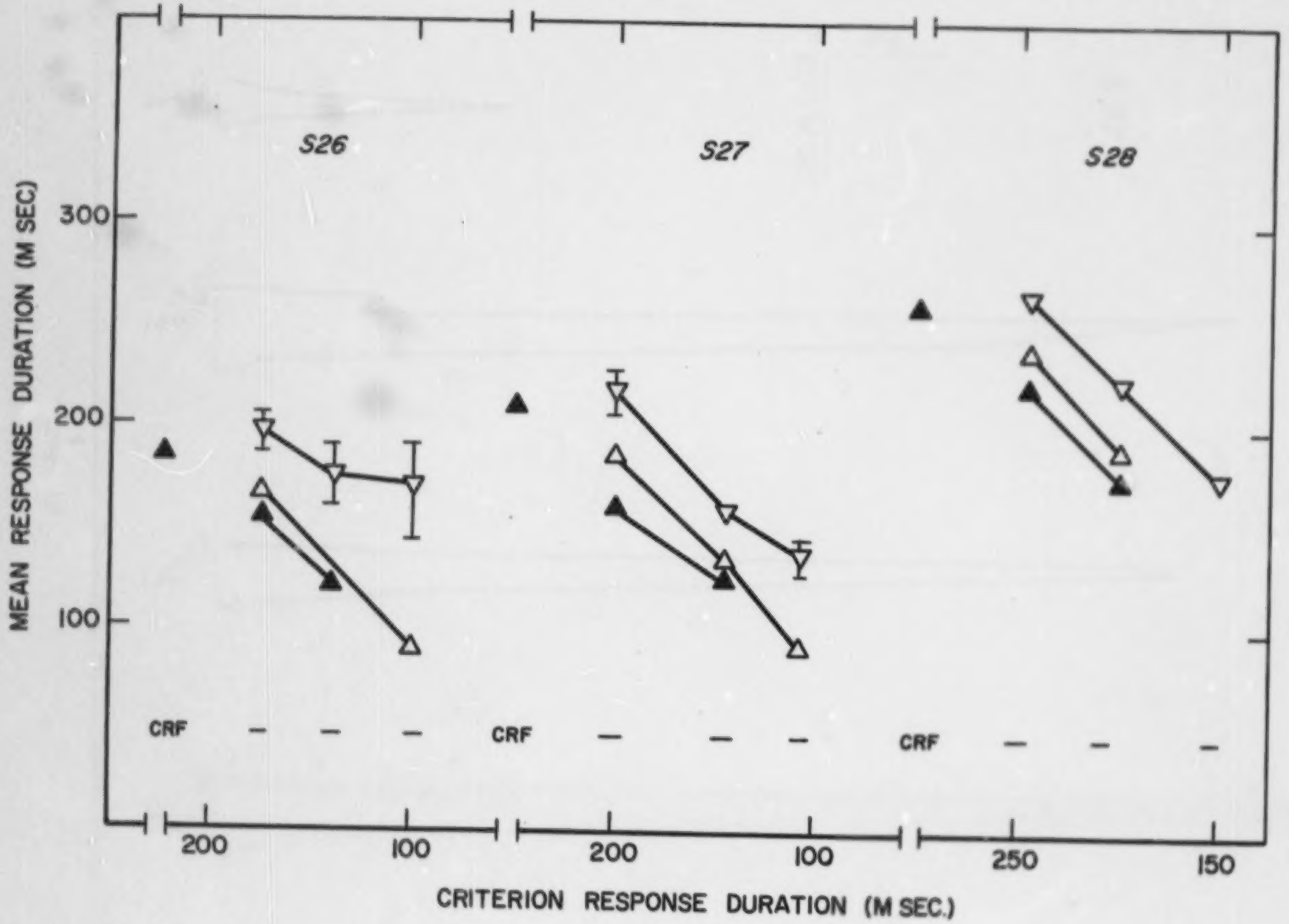

S26
S27
S28
MEAN RESPONSE DURATION (M SEC)
300
200
100
CRF
CRF
CRF
200
100
200
100
250
150
CRITERION RESPONSE DURATION (M SEC.)

Fig. 15

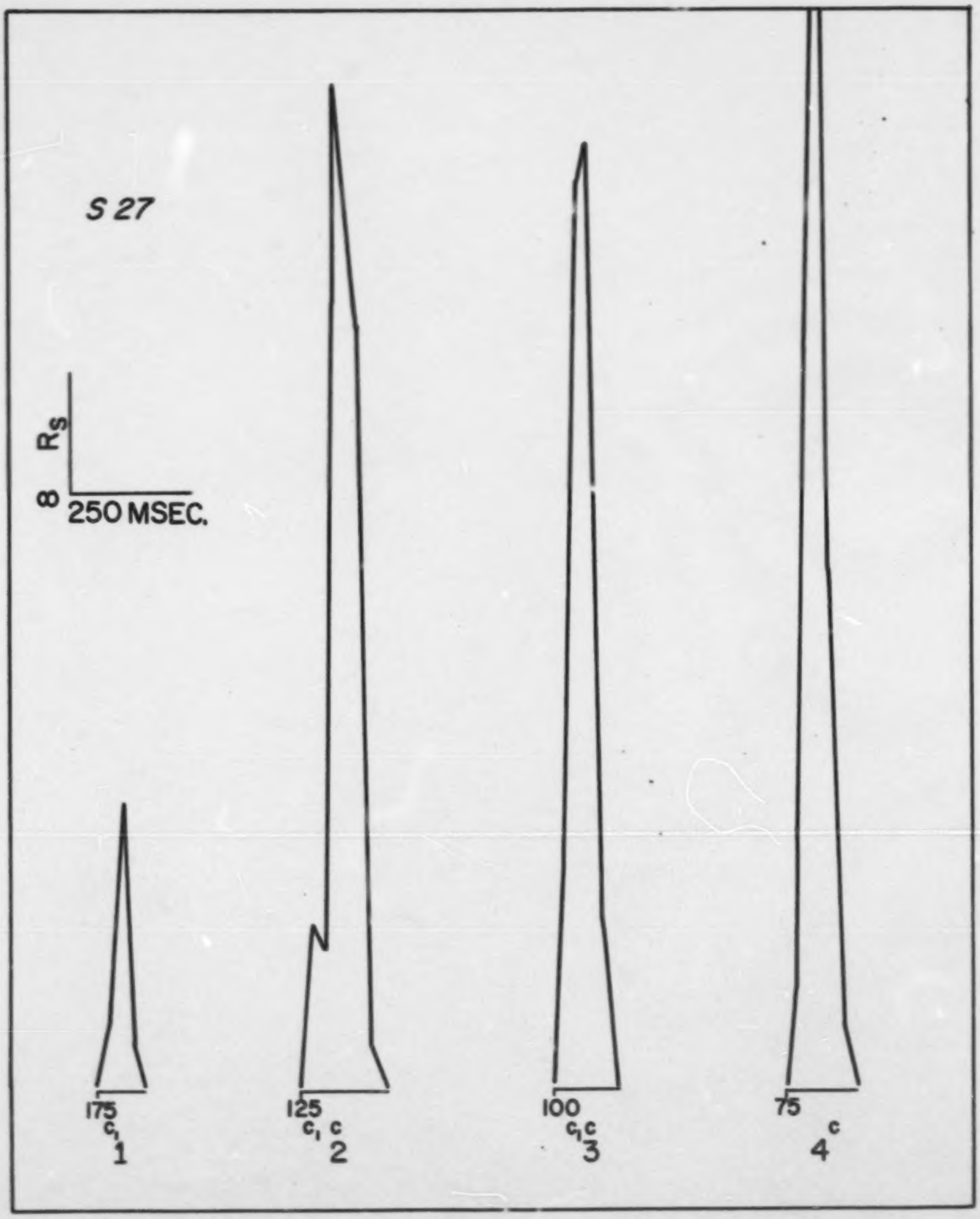


Fig. 16

METRIC PROPERTIES OF MULTIDIMENSIONAL STIMULUS GENERALIZATION

D. V. Cross
The University of Michigan

Metric Properties of Multidimensional Stimulus Generalization

D. V. Cross

Behavior Analysis Laboratory

University of Michigan

Introduction

Often the stimuli which set the occasion for differential responding differ on two or more physical dimensions and may even involve different sensory modalities. How does the behavior evoked in these situations compare with discriminative behavior that is under the control of stimulus properties which vary along one physical dimension only? What are the rules that govern how stimulus differences along two or more dimensions combine to exert control over responding? These questions may be raised with regard to a variety of experimental problems: whether we are concerned with the course of development of stimulus control, as in discrimination-learning experiments; with the specificity of stimulus control, as in generalization experiments; or with the relation between specific properties of stimuli and presumptive measures of sensory magnitude, as in psychophysical experiments. Unfortunately, experiments investigating the behavioral consequences of varying a discriminative stimulus along more than one dimension are rare. Relatively little is known about how unidimensional stimulus differences combine to effect multidimensional stimulus control. Although it is true that the methods of multidimensional psychophysics rest upon certain highly specific assumptions about these rules of combination, the adoption of these rules has followed the dictates of mathematical convenience rather than those of conclusive experimental evidence.

Behavioral control and stimulus dimensionality

The first, perhaps most elementary, point of concern is the question of whether or not stimuli differing on several dimensions exercise more behavioral control than do stimuli which differ on one dimension alone. There exists contradictory evidence in this regard, if we accept as relevant the results of all experiments in which the fact or degree of stimulus control is the primary variable under investigation. These experiments include studies of stimulus generalization, in which a response is brought under the control of a particular stimulus complex and then tested in the presence of other stimuli. Included also are conditioning experiments and discrimination-learning experiments, in which the dependent variable is either the number of stimulus presentations, or trials, required to establish control, or it is the degree of control established, as measured by the amount of conditioning, or the number of correct discriminations obtained. For example, Miller (1939) found that a combined auditory and visual stimulus (a drop in pitch of a continuous tone, and the movement of a pointer), paired with electric shock to the cheek, resulted in faster and stronger conditioning of an eyelid response in humans than did either stimulus property used alone. Similarly, Eninger (1952) found that rats formed a discrimination much more rapidly when auditory and visual cues were combined (tone vs no-tone, and white vs black) than when either of these sets of cues was presented alone. In a generalization study, Fink & Patton (1953) controlled certain visual, auditory, and tactile characteristics of the stimulus situation in which rats learned to drink water from a tube. After establishing a stable baseline of drinking rate, the investigators altered one or more of the stimulus properties under their control. They found that any change led to a decrement in drinking: light changes were most effective, sound changes were moderately effective, and tactual changes were least effective in modifying the strength of the learned drinking response. More important, however, they found that a change in any two stimulus properties resulted in a greater response decrement than a change in any single property; when all three stimuli were changed, maximum response decrement occurred. White (1958) trained children to pull a handle to get marbles, using a card of a certain hue and saturation (Munsell 10GY 8/6) as a discriminative stimulus. Then, for different groups of subjects, he changed the hue of the stimulus card, its saturation, or both, in a series of unreinforced test trials. There was greater generalization decrement to test stimuli which differed from the training stimulus in both dimensions than to those which differed in either of the dimensions alone. That is, the properties of hue and saturation combined to exert more control over differential responding than either property exerted alone. A generalization suggested by these results is that every stimulus property that happens to be correlated with the contingencies of reinforcement requiring differential responding acquires some degree of stimulus control; when stimulus properties are presented in combination, their effects combine in some manner to exert an even greater degree of control.

On the other hand, it has been found, in some discrimination experiments, that stimuli differing in two properties are discriminated with no greater ease or efficiency than stimuli differing in the single one of these properties which is most readily discriminated. In other words, the degree of behavioral control exerted by the several properties of a compound stimulus is no greater than that exercised by the component property which, by itself, is most effective. Harlow (1945) trained monkeys to respond discriminatively to stimuli differing in color only, form only, or both. Stimuli differing in both color and form, or in color alone, were discriminated equally well, although both were discriminated more readily than form. This result was also found by Warren (1953) who tested monkeys on discrimination problems involving stimuli which differed in color, form, size or in the four combinations of these properties taken two or three at a time. Stimuli which differed in color and some other property were little, if any, more discriminable than those differing in color alone. In a later experiment, Warren (1954) trained monkeys to discriminate pairs of stimuli which differed simultaneously in color, form, and size. After establishing stimulus control, the number of discriminative cues was reduced by eliminating stimulus differences in one or two dimensions. In each instance, performance scores were depressed. Discriminative control was almost completely lost when color differences were eliminated, regardless of which other property or combination of properties remained. Some control was lost when form and size differences were both eliminated, and the smallest decrements in performance occurred when only form or size differences were eliminated, size being the least effective.

We may conclude that the findings of some studies imply a distribution of behavioral control over the several stimulus properties, while others suggest that a particular subset--or only one--of the stimulus properties is dominant, i.e., controls differential responding.

Metric properties of stimulus control

In order to make quantitative predictions concerning the control of responding by complex stimuli we require equations expressing the relation between changes in one or more stimulus properties and the correlated changes in the properties of responding. It is generally the case that the difference in degree of control exercised by two stimuli separated on a single physical continuum is not an invariant function of their physical difference but varies markedly from one location on the continuum to another. To achieve the desired translation invariance, the behavioral quantities measured in relation to changes in the physical stimuli are used to induce a transformation on the stimulus continuum, resulting in a scale which is more useful in predicting differential responding than the original physical scale. When stimuli differ along several physical dimensions, however, the scaling problem becomes more complicated. For, in addition to the unidimensional scales themselves, special rules are needed for specifying exactly what combinations of differences on these separate scales will define a stimulus that exerts the same behavioral control as that exercised by a stimulus defined by one scale alone. There are three major points of view concerning these rules of combination as they pertain to multidimensional stimulus generalization. Two-dimensional models of these views are presented graphically in Fig. 1. The two stimulus dimensions are labelled S1 and S2, respectively, and their point of intersection represents the training stimulus. Degree of generalization is depicted as a third dimension, labelled R, orthogonal to the other two. The stimulus dimensions are assumed to be appropriately scaled so that the generalization gradients for the separate dimensions are linear and of equal slope. The task of the models is to predict the height of the generalization surface at any point in the four quadrants of the stimulus plane. Theoretical justification for the first model, the excitation model, comes from Pavlov, and later Hull, who conceived of generalization as resulting immediately as a spread of excitation in all directions from the training stimulus. According to this view, prior differential training along specific stimulus dimensions is not necessary to obtain gradients of generalization. Thus, for a generalization surface of given dimensionality, one set of dimensions or frame of reference is as good as any other in predicting generalization behavior. Although he never stated it explicitly, Hull implies that a generalization surface for two dimensions can be generated by rotating a unidimensional gradient around its peak, resulting in a surface which is conical in shape, as depicted in the first panel of Fig. 1. A contrasting view, proposed by Lashley and Wade (1946) and by Guttman (1956), is that generalization is a performance phenomenon reflecting only the discriminative control exercised by stimuli varying along dimensions previously established for the organism by differential training. This view predicts that the generalization decrement for any stimulus varying in two dimensions is the sum of the decrements to stimuli changed in each dimension alone. The generalization surface has the appearance of a pyramid with its four corners on the stimulus axes. This discrimination model is depicted in the second panel of Fig. 1. The dimensional dominance model is suggested by the results of the experiments, discussed in the prior section of this article, in which some one stimulus property appeared to be "dominant" in exerting behavioral control. Whereas the discrimination model requires that an organisms' behavior is under the unselective control of all discriminable changes in the stimulus, the dimensional dominance model states that stimulus control is selective -- exerted only by the dimension, previously established for the organism by differential training, along which the most discriminable stimulus change occurred. Theoretical background for this model may be found in Lashley's "principle of dominant organization." Lashley (1942) maintains that the mechanism of nervous integration is such that, when any complex of stimuli arouses nervous activity, that activity is immediately organized and certain stimulus properties become dominant for reaction while others become ineffective. Of course, if only one stimulus dimension were to show differential control, we would characterize other dimensions as irrelevant rather than as dimensions of generalization or discrimination. In this case, the generalization surface would collapse to a single dimension. The generalization surface predicted by the dimensional dominance model is also a pyramid with its corners at 45° to the stimulus axes.

If we look at successive sections through the generalization surfaces depicted in Fig. 1, sections taken horizontal to the respective stimulus planes, sets of concentric contours are revealed. These contours are circular for the excitation model, diamond shaped (rotated squares) for the discrimination model, and square for the dominance model. Each contour corresponds to a prescribed amount of generalization decrement and describes the locus of equally effective stimuli. That is, all stimuli on a contour are equivalent with respect to the amount of control exercised over differential responding, and thus equally substitutable one for another to produce a given generalization decrement. The question of the shape of these contours turns out to be equivalent to the question of just what metric is appropriate for a given set of measures. The shape of the contours uniquely determines, and is determined by, the structure of the underlying metric space. Thus, the assumption of the excitation model that all contours are circular, is equivalent to the very special assumption that distances, i.e., the amount of generalization decrement, or the separation between the training stimulus and the test stimulus on the underlying scale, are Euclidean. The concept of "distance", applied to behavioral measures, is related to its use in ordinary spatial representations. Technically, any set of numbers satisfying the metric axioms can be called a distance. In particular, a set R of behavioral measures is a metric space, if a real-valued function $D(x,y)(x,y \in R)$, called the distance between x, y is defined for all pairs of elements x, y of R, so as to satisfy the following axioms:

M_1 $D(x,y) = D(y,x)$ (Symmetry)

M_2 $D(x,y) > 0$ if $x \neq y$, and $D(x,x) = 0$ (Positivity)

M_3 $D(x,y) \leq D(x,z) + D(z,y)$; $x,y,z \in R$ (Triangle inequality)

It is a strictly empirical question whether or not our measurement operations result in quantities conforming to these axioms.

In Euclidean space, the distance between two points is given by the Pythagorean formula, i.e., is equal to the square root of the sum of squared differences between the projections of the two points on each of their coordinate axes. This is the basic assumption underlying most multidimensional scaling models. Multidimensional scaling models based on the Eudlidean metric have been used in several studies involving color stimuli. In general, the model fits the behavioral data well. Shepard (1958) constructed a model of this kind applicable to stimulus-response confusion data, typically obtained in paired-associate (identification learning) experiments. The model is based on the assumption that stimulus generalization, in terms of probability of a correct identification, is an exponential decay function of distance in Euclidean space. The model was applied to a matrix of confusion probabilities obtained from subjects required to identify each of nine Munsell colors varying in brightness (value) and saturation (chroma). Approximately 99% of the original variance in the data was accounted for by the model. On the other hand, in certain experimental settings, model of stimulus generalization based on a non-Euclidean metric has proven to be a more appropriate description of the findings. The non-Euclidean metric associated with the discrimination model provides a particularly simple rule for combining distances. Namely, the distance between two points is equal to the sum of the differences between the projections of the two points on each of their coordinate axes. This is often called the "city-block" metric because the distance between two points is the total distance that must be traversed in a north-south direction plus the distance that must be traversed along the east-west direction. Attneave (1950) found that measures both of judged similarity and of frequency of confusion among simple geometrical figures (triangles or parallelograms), varying along the dimensions of size and either brightness or shape (tilt), conformed more to the city-block metric than to the Euclidean Metric.

In an attempt to discover what rule-of-combination applies in multidimensional stimulus generalization, Butter (1963) trained pigeons to peck at a key which was illuminated by a narrow vertical strip of monochromatic (550 mm) light. The birds were tested under extinction conditions for generalization to stimuli varied in the wavelength dimension, in the angular orientation (tilt) dimension, and in both dimensions. There was greater generalization decrement to stimuli changed in both dimensions than to changes in either dimension alone. Analysis of variance performed on the generalization measures revealed a significant interaction between the two stimulus dimensions. The source of this interaction was revealed in a graph depicting generalization as a function of stimulus changes along one dimension, with levels of the other dimension as the parameter. The gradients in the family of curves thus described were not parallel but appeared to be multiples of each other. This suggested that relative generalization to a stimulus that is changed in two dimensions simultaneously equals the product of the values of relative generalization obtained when the two changes are made independently, i.e., along single dimensions alone. If we denote by R(i,j) the number of responses emitted in the presence of a stimulus different from the training stimulus by i units along one dimension and by j units along the second dimension, then

$$\frac{R(i,j)}{R(0,0)} = \frac{R(0,j)}{R(0,0)} \cdot \frac{R(i,0)}{R(0,0)},$$

where R(0,0) denotes the number of responses evoked by the training stimulus. This "multiplicative rule" was tested against the obtained data and was found to predict the number of responses to generalized stimuli with a remarkably high degree of accuracy. The investigator concluded that this result was not consistent with the discrimination model. On the other hand, it is proper to point out that when ratios multiply, logarithmic differences summate. The significant interaction between dimensions observed in this study was not intrinsic to the dimensions but was essentially an artifact of the scale of measurement employed. If the analysis had been performed on logarithmic values, the interaction term would have vanished and the treatment effects would have become additive. The appropriateness of a logarithmic transformation on these data is evidenced by the observed dependence of response variability on response magnitude; scaling the dependent variable to render measurement equally precise at all levels of the scale (the achievement of variance homogenity) meets a desirable measurement criterion. Thus, Butter's results may be interpreted as favoring the discrimination model.

The metric for the dimensional dominance model requires that the distance between two points is simply the greatest of the distances separating their projections on each of the coordinate axes. This metric has never been formally employed in any multidimensional scaling model although the rule of combination inducing the metric was indirectly employed by Eriksen and Hake (1955) to predict the discriminability of stimuli varying in two or three dimensions from the discriminability of stimuli varying along only one dimension at a time. Their subjects were trained to identify each stimulus, in series of up to 20 stimuli varying in either size, hue, brightness, or the four combinations of these taken two or three at a time, by assigning different numbers to each stimulus. The probability of assigning a particular number, during testing, to a stimulus varying in two or three dimensions, was predicted from the response probabilities obtained from the unidimensional series. The predictions were based on the assumption that a stimulus of a given hue, size, and brightness evokes either response $\underline{i}$ under the control of hue, response $\underline{j}$ under the control of size, or response $\underline{k}$ under the control of brightness. The response with the greatest probability was assumed to be the response that would be emitted. Predictions of response probabilities were based on the assumption of independence of events, and involved computation of the joint probability associated with each event. The probability that a particular response $\underline{i}$ would be emitted to a compound stimulus was predicted by summing the joint probabilities of all events in which response $\underline{i}$ was most probable. The predictions fit the empirical probabilities notably well.

Each of the metrics just discussed has different implications for interpreting the quantitative properties of behavior under complex stimulus control. These do not exhaust all the possible metric spaces of potential psychological interest. In fact, consistent geometrics have been developed for more general classes of metric spaces of which the three described are only special cases. In all cases where the obtained contours are centrally symmetric, convex, and of the same shape, there exists a distance function which will map the underlying measures into a metric space. Of particular interest is the class of metric spaces generated by the Minkowski distance function which defines the distance between the origin 0 and an arbitrary point P in two-dimensional space as

$$d_r\ (OP) = (|x|^r + |y|^r)^{1/r}$$

for some fixed value of $r \geq 1$. In application to stimulus generalization, the training stimulus may be identified as the origin of the space, a test stimulus as the point P, and x and y as the amount of generalization decrement along the respective stimulus dimensions with respect to which P may vary. The "unit circles", i.e., the locus of stimulus points at a distance 1 from the origin is given by

$$|x|^r + |y|^r = 1.$$

The shape of the unit circle when graphed relative to cartesian coordinates depends on the value of r (Fig. 2). When $r = 1$ we have the city-block "unit circle"; for $r = 2$, we have the Euclidean "unit circle"; and as r approaches infinity, d_r (OP) approaches the distance

$$d_r\ (OP) = \max\{|x|,\ |y|\},$$

and the "unit circle" for $\max\{|x|, |y|\} = 1$ is given by the square with sides

$$|x| = 1, \qquad 0 \leq |y| \leq 1$$
$$|y| = 1, \qquad 0 \leq |x| \leq 1.$$

This is the metric implied by the dimensional dominance model discussed above. It can be shown (Beckenbach & Bellman, 1961, pp. 103-106) that Minkowski distance satisfies the metric axioms described earlier as well as certain properties deduced from them; it is translation invariant, symmetric, always positive, and it satisfies the conditions of triangle inequality and homogeneity.

Shepard (1962) performed four ingeniously designed experiments to determine the particular Minkowski metric that describes how differences in size and angular orientation combine to exert control over differential responding in various tasks. The stimuli employed were circles with single radial lines (spokes) drawn at different angles. Shepard analyzed the frequencies with which stimuli varying in size, inclination, or both, were matched with a standard stimulus and the frequencies with which stimuli were actually confused during identification learning. When the results were based on data pooled over all subjects, they could not be accounted for by any static metric solution. The obtained isosimilarity contours were concave, indicating a violation of the condition of triangle inequality. Shepard concluded that this anomalous result was due to the indiscriminate pooling of data from subjects who were attending to different aspects of the stimuli. The subjects were divided, therefore, into two groups; the data from subjects whose behavior appeared to be under the control of both stimulus properties were analyzed separately from the data of subjects whose behavior appeared to be dominantly under the control of one or the other stimulus property. The subjects in the latter group were described as adopting a "matching strategy" which essentially involved "finding the best match along one of the two dimensions." Although the point was not made by Shepard, the data from these subjects are consistent with the Minkowski metric characteristic of the dimensional dominance model. Data from the former group, presumed to have attended both dimensions, suggested that the isosimilarity contours were roughly four-cornered, indicating the appropriateness of a Minkowski metric close to the "city-block" variety. The results of the identification-learning experiment also tended to favor the "city-block" metric, even for individual subjects.

The above findings suggest that the different properties of a multidimensional stimulus may be selectively effective in controlling differential responding. Consider two stimuli differing with respect to properties X and Y. One stimulus might be the training stimulus in a generalization experiment and the other might be a particular test stimulus. If both stimuli evoke the same response, then neither X nor Y exercise discriminative control. If the stimuli evoke differential responding, however, then this can be attributed to fact that discriminative control is exercised by either the difference with respect to property X alone, or to Y alone, or to both. If it is assumed that these four events are mutually exclusive and exhaustive, then their probabilities sum to 1, that is,

$$(1) \quad P(XY) + P(X\bar{Y}) + P(\bar{X}Y) + P(\overline{XY}) = 1.00$$

where:

$P(XY) \equiv$ the probability that no differential control is exerted

$P(X\bar{Y}) \equiv$ the probability that differential control is exerted only by Y

$P(\bar{X}Y) \equiv$ the probability that differential control is exerted only by X

$P(\overline{XY}) \equiv$ the probability that differential control is exerted by both X and Y simultaneously.

These probabilities might also be interpreted as corresponding to various states of attention or the presence of particular observing or orienting responses. For example, $P(\bar{X}Y)$ might be interpreted as the probability that the subject attends or observes or orients to only the difference in *X*.

The probability, $P(XY)$, that a stimulus changed with respect to both *X* and *Y* will evoke the response appropriate to the training stimulus, is given by

$$(2)\quad P(XY) = 1 - [P(\bar{X}) + P(\bar{Y}) - P(\overline{XY})]$$

where:

$P(\bar{X}) \equiv$ the marginal probability that *X* will exert differential control
$P(\bar{Y}) \equiv$ the marginal probability that *Y* will exert differential control.

If these probabilities are independent, then $P(\overline{XY}) = P(\bar{X})\,P(\bar{Y})$ and equation (2) simplifies to

$$(3)\ P(XY) = P(X)\,P(Y).$$

Thus, response probabilities to stimuli changed in two dimensions can be predicted by multiplying response probabilities to stimuli changed in each dimension alone. The isosimilarity contours generated by this rule of combination change drastically from one probability level to another (see Fig. 3). They appear to span the entire range of the Minkowski distance function from $r = 1$ to $r \rightarrow \infty$. The contours generated are not Minkowski circles, however, but are rectangular hyperbolic arcs. It follows from the present analysis that the shape of the isosimilarity contours depends upon the distance separating the variable stimulus array from the standard. In the series of stimuli employed by Shepard, as initially scaled, the stimulus arrays were all equi-distant from the standard. Arrays either closer to or more distant from the standard might have yielded different results.

The discussion up to this point has been based on the assumption that the probabilities defined for equation (2) are independent. Suppose, instead, that these probabilities are not independent, but conditionally dependent. Two cases of particular interest are encountered: when the conditional probabilities are either 0 or 1.

Case I.

$$P(\overline{X}|\overline{Y}) = P(\overline{Y}|\overline{X}) = 0 \Longleftrightarrow P(\overline{X}\,\overline{Y}) = 0, \text{ then}$$

$$P(XY) = 1 - [P(\overline{X}) + P(\overline{Y}) - 0] \text{ or,}$$

$$1 - P(XY) = [1 - P(X)] + [1 - P(Y)]. \tag{4}$$

This is precisely the "city-block" metric which predicts that the generalization decrement to a stimulus changed with respect to two properties equals the sum of the decrements to either property alone. The interesting fact is that this is deduced from the assumption that $P(\overline{X}\,\overline{Y}) = 0$, i.e., that differential responding is never under the control of both properties simultaneously.

According to Shepard, a static metric solution to multidimensional discrimination data is not possible when a subject's attention shifts from one stimulus property to another on different presentations of the same physical stimulus. The above probability analysis of complex stimulus control, however, deduces the "city-block" rule of combination for response probabilities from the explicit assumption that the subject's responding is never under the control of both stimulus properties simultaneously.

Case II.

$$\text{If } P(\bar{X}|\bar{Y}) = 1 \text{ when } P(\bar{Y}) < P(\bar{X}), \text{ then}$$

$$P(XY) = 1 - [P(\bar{X}) + P(\bar{Y}) - P(\bar{X}|\bar{Y})\, P(\bar{Y})] \text{ and}$$

$$P(XY) = 1 - P(\bar{X}) = P(X) \text{ or}$$

$$(5)\quad 1 - P(XY) = \max\,[P(\bar{X}),\, P(\bar{Y})].$$

Equation (5) is equivalent to the metric implied by the *dimensional dominance model*.

The present study was designed to examine multidimensional stimulus control of responding when the dominance of the stimulus dimensions was under experimental control. The pattern of reinforcement for discriminative responding by the subject to members of the two-dimensional array of auditory stimuli was contingent on a higher-order visual discrimination. For one value of the visual cue, differential responding with respect to one dimension of the stimulus array was reinforced; for a second value of this cue, differential responding with respect to the orthogonal auditory dimension was reinforced. In generalization testing, the two values of the visual cue were presented singly and in combination. Under these conditions, we examine the metric properties of stimulus generalization; in particular, we consider the appropriateness of the "city-block" metric when differential responding to one or the other stimulus dimension is independently controlled.

Method

Subjects

Twenty male, undergraduate students, who had no previous experience in generalization or discrimination experiments, served as subjects (*S*s).

Stimuli and apparatus

The stimuli were sinusoidal tones varying in both fundamental frequency and the rate at which they were frequency-modulated ± 20 cps. The tones were generated by a beat frequency oscillator (Bruel and Kjaer 1014) and pre-recorded on magnetic tape (Ampex 351-2U) at four different levels of fundamental frequency: 500, 600, 750, and 900 cps, and of modulation rate: 2, 4, 8, and 16 sweeps/sec (sps). All combinations of fundamental frequency with modulation rate were used, yielding a set of 16 stimuli. The experimental space was a sound-attenuating room, containing a rack-mounted panel and shelf. On the front of the panel were two pilot lamps of different color, amber and blue, and a tray for receiving pennies. An electrically operated coin dispenser, obtained from a vending machine and modified for experimental use, was mounted behind the panel. A response key (Poucel Electronics) that could be pressed either to the left or to the right was mounted on the shelf. The chamber was dark except when one or both of the pilot lamps was lit. The control equipment was in an adjoining room. The stimuli were reproduced by the tape recorder and presented to S through calibrated earphones (Grason-Stadler TDH-39). Interposed between the tape recorder and the earphones was an electronic switch (Grason-Stadler 829D) that closed upon presentation of the stimulus and opened when the response key was pressed thus terminating the stimulus. The electronic switch, as well as occasions for reinforcement and other experimental events were all controlled by voice-operated relays (Miratel), which were operated, in turn, by coding tones synchronized with the stimuli and pre-recorded on a second track of the magnetic tape. Response latencies were measured by recording, on a second tape recorder (Ampex 601) 1000-cps tones passed through a second channel of the electronic switch. The duration of these tones corresponded exactly to the time interval from the onset of the stimulus to its termination by a response. On subsequent play-back these durations were measured by a frequency counter (Hewlett-Packard) and automatically punched on IBM cards for analysis.

Procedure

The Ss were seated individually in the experimental chamber and the following instructions were read.

"You can earn money in this experiment if you can learn to correctly respond to certain properties of sounds that you will hear in these earphones. You will respond by pressing this key either to the left or to right. If your response is correct, a penny will be dispensed to you — if not, you get nothing. In either case, the response you make, right or wrong, will terminate the stimulus. The basis for a correct discrimination will depend upon which of these two lights is on at the time. The faster you learn and the fewer errors you make, the more money you will earn."

Discrimination training. Three stimuli were presented, five times each, in 10 randomized sequences for a total of 150 stimulus presentations. During five sequences the amber light was on and during the other five sequences, the blue light was on. (The order was randomized.) Under either light condition, a 500 cps/2 sps signal was an S^D for the response of pressing the key to the left, R_1. A 500 cps/16 sps signal was also an S^D for R_1 when the amber light was lit, but when the blue light was on, it was an S^D for a key press to the right, R_2. A 900 cps/2 sps signal was an S^D for R_2 when the amber light was on, and for R_1 when the blue light was on. The stimuli were presented at 10-sec intervals for a maximum duration of 5 sec on each presentation. If at the end of 5 sec S did not respond, the stimulus terminated and the reinforcement mechanism was inoperative. Otherwise the stimulus was terminated by the response, whether it was correct or not. If correct, a penny was dispensed and a bell, with a sound similar to that of a cash register, was rung. There was a 30-sec time out between each stimulus sequence.

Generalization testing. At the completion of training, the experimenter re-entered the chamber and told S that, for the remainder of the experiment, the coin dispenser and bell would be disconnected. In addition, S was instructed to respond to each stimulus. All Ss were presented with 10 random permutations of the 15 stimuli in sequence, with a 30-sec time out between each sequence. For ten Ss the amber light was on for half of the sequences and the blue light was on for the other half. For the other ten Ss, two of the first four stimulus sequences were presented with the amber light on and the other two with the blue light on. For the remaining six stimulus sequences, both the amber and the blue were on.

Results

Formation of a correct discrimination was rapid for all Ss. Very few errors were made after the fourth stimulus sequence. Figure 4 shows that discriminative control was established over differential responding to one or the other of the dimensions of the auditory stimuli. Only the R_1 gradients are represented because the R_2 gradients are complementary functions. The functions depicted represent the over-all discriminative control exercised by each dimension under the two light conditions. When the amber light was on, discriminations were clearly based on changes in fundamental frequency only, and when the blue light was on, modulation rate exercised exclusive control. There was virtually no interaction between the levels of either dimension under the two conditions. The generalization gradients for the controlling dimensions in both conditions coincided at all levels of the irrelevant dimension, and the generalization gradients along the irrelevant dimension, for different levels of the dominant dimension were parallel. Having established the fact of selective stimulus control, we may pool the results obtained under the two light conditions and compare the "city-block" rule of combination with the multiplicative and dominance rules of combinations in describing the generalization gradients. The primary gradients along the dimension of fundamental frequency, with modulation rate fixed at 2 sps, and along the dimension of modulation rate, with fundamental frequency set at 500 cps, were used to predict the decrement in the number of R_1 responses (relative to the number of responses emitted to the training stimulus) at each of the other nine stimuli. Figure 5 presents a plot of obtained decrements vs. predicted decrements for the "city-block" rule (diamonds), the dominance rule (squares) and the multiplicative rule (circles). The straight line drawn at 45° to the axes represents perfect agreement between obtained and predicted generalization decrements. It will be seen that, of the three, the "city-block" rule is the best predictor. Only two out of the nine "city-block" predictions deviate from the ideal line by more than three responses, as compared to six out of nine for the multiplicative rule and all nine out of nine for the dominance rule.

Response latency provides a second and highly sensitive index of discriminative control (Fig. 6). Latencies were not plotted for four stimuli because either no responses or only one or two responses were emitted. The latency gradients along the dominant dimension under each light condition tend to be steeper than the latency gradients along the irrelevant dimension. This shows up particularly well for the R_2 latencies.

Figure 7 permits a more detailed examination of the R_2 latency gradients. The gradients in Fig. 7 represent response latencies at each level of one dimension with levels of the other dimension as parameters. These gradients show surprising uniformity.

Figure 8 presents the results of testing with both the amber and the blue lights on simultaneously. It is apparent that both dimensions exercise behavioral control with respect to overall results. Latencies show a general decreasing trend which, except at three data points, is fairly systematic. The unfilled circles in the latency gradients (top graphs) indicate that these latencies are based on less than five responses and thus unreliably estimated. The number of responses entering into latency estimates for all the points is given directly by the gradients in the lower half of Fig. 8. The diamonds in this figure, appropriately enough, represent the number of responses predicted by the "city-block" rule of combination. The prediction was based on the number of responses to those stimuli identified by the unfilled circles (bottom graphs). This set of data points corresponds to the R_2 generalization gradient with respect to the common S^D (500 cps/2 sps). Except for one point (at 600 cps/16 sps) the predicted quantities fit the obtained rather well, indicating the appropriateness of the metric for these data.

Summary

Twenty human Ss were employed in a study of multidimensional stimulus generalization. The pattern of reinforcement for discriminative responding (key-press) by S to the 16 members of a two-dimensional array of auditory stimuli, varying in center frequency and modulation rate, depended on a higher-order visual discrimination. Differential responding along each dimension during discrimination training came under the control of the correlated light condition (amber or blue). In generalization testing, each or both of the lights was presented during the several series of auditory stimuli, and the conditional frequency and latency of responding was observed.

Generalization gradients of response frequency and latency were found to be steep along the dimension correlated with the concurrent visual cue, and flat along the orthogonal dimension. When both visual cues were presented concurrently, both dimensions of the auditory stimuli exercised control over differential responding.

Among the possible Minkowski metrics that may be employed to predict the shape of the generalization surface, based on the pair of gradients obtained for the two dimensions at the levels which intersect at the training stimulus, the "city-block" metric, which corresponds to the "discrimination model" of multidimensional stimulus generalization, was found to be most accurate.

References

Attneave, F. Dimensions of similarity. *Amer*. *J*. *Psychol*., 1950, 63, 516-556.

Beckenbach, E. & Bellman, R. *An introduction to inequalities*. New York: Random House, 1961.

Butter, C. M. Stimulus generalization along one and two dimensions in pigeons. *J*. *exp*. *Psychol*., 1963, 65, 339-346.

Eninger, M. U. Habit summation in a selective learning problem. *J*. *comp*. *physiol*. *Psychol*., 1952, 45, 604-608.

Eriksen, C. W. & Hake, H. W. Multidimensional stimulus differences and accuracy of discrimination. *J*. *exp*. *Psychol*., 1955, 50, 153-160.

Fink, J. B. & Patton, R. M. Decrement of a learned drinking response accompanying changes in several stimulus characteristics. *J*. *comp*. *physiol*. *Psychol*., 1953, 46, 23-27.

Guttman, N. The pigeon and the spectrum and other perplexities. *Psychol*. *Rep*., 1956, 2, 449-460.

Harlow, H. F. Studies in discrimination learning by monkeys: VI Discrimination between stimuli differing in both color and form, only in color, and only in form. *J*. *gen*. *Psychol*., 1945, 33, 225-235.

Lashley, K. S. An examination of the "continuity theory" as applied to discriminative learning. *J*. *gen*. *Psychol*., 1942, 26, 241-265.

Lashley, K. S. & Wade, M. The Pavlovian theory of generalization. *Psychol*. *Rev*., 1946, 53, 72-87.

Miller, J. The rate of conditioning of human subjects to single and multiple conditioned stimuli. *J*. *gen*. *Psychol*., 1932, 20, 399-408.

Shepard, R. N. Stimulus and response generalization: tests of a model relating generalization to distance in psychological space. J. exp. Psychol., 1958, 55, 509-523.

Shepard, R. N. Attention and the metric structure of the stimulus space. Unpublished report, Bell Telephone Laboratories, Murray Hill, New Jersey.

Warren, J. M. Additivity of cues in a visual pattern discrimination by monkeys. J. comp. physiol. Psychol., 1953, 46, 484-486.

Warren, J. M. Perceptual dominance in discrimination learning by monkeys. J. comp. physiol. Psychol., 1954, 47, 290-293.

White, S. W. Generalization of an instrumental response with variations in two attributes of the CS. J. exp. Psychol., 1958, 56, 339-343.

Figure Captions

Fig. 1. Possible forms of the generalization surface in two dimensions.

Fig. 2. Euclidean graphs of Minkowski "unit circles."

Fig. 3. Generalization contours generated by the multiplicative rule.

Fig. 4. Auditory generalization under visual control. Each point represents the total number of R_1 responses emitted (of a possible 280) at each level of a given dimension, summed over levels of the orthogonal dimension, as a function of color of the visual stimulus. During training, the amber light was a discriminative stimulus controlling differential responding to differences in fundamental frequency and the blue light controlled differential responding to differences in modulation rate.

Fig. 5. Comparison of three rules of combination in the prediction of generalization decrements. Generalization decrement measures, in terms of the difference between the number of R_1 responses to the training stimulus and test stimuli on the primary generalization dimensions (intersecting at the training stimulus) were used to predict the decrement in responding to each of the remaining nine stimuli. Predictions based on the "city-block" rule of combination are represented by diamonds, the dominance rule by squares, and the multiplicative rule by circles.

Fig. 6. Generalization gradients of response latency. Both R_1 latency gradients (top graphs) and R_2 latency gradients (bottom graphs) are shown as a function of light condition during testing. Each point represents the geometric mean latency of all responses to a single level of a given dimension. Points were not plotted if the corresponding latencies were based on too small a sample size (two or less determinations).

Fig. 7. Generalization gradients of response latency. The geometric mean latency of all R_2 responses to a single stimulus are represented as a function of one dimension with levels of the other as parameters, and as a function of the color of the controlling visual stimulus.

Fig. 8. Generalization under the combined light conditions. Both R_2 latency (top graphs) and frequency (bottom graphs) gradients for each dimension, with levels of the other dimension as parameters, are shown as a function of both visual stimuli presented together. The unfilled circles in the top graphs represent geometric mean latencies computed on five responses or less. The unfilled circles in the lower graphs represent the number of responses emitted to stimuli on the primary dimensions. These quantities were used to predict generalized responding to the other stimuli, represented by filled circles. The predictions for each family of gradients are depicted by the diamonds connected by broken lines.

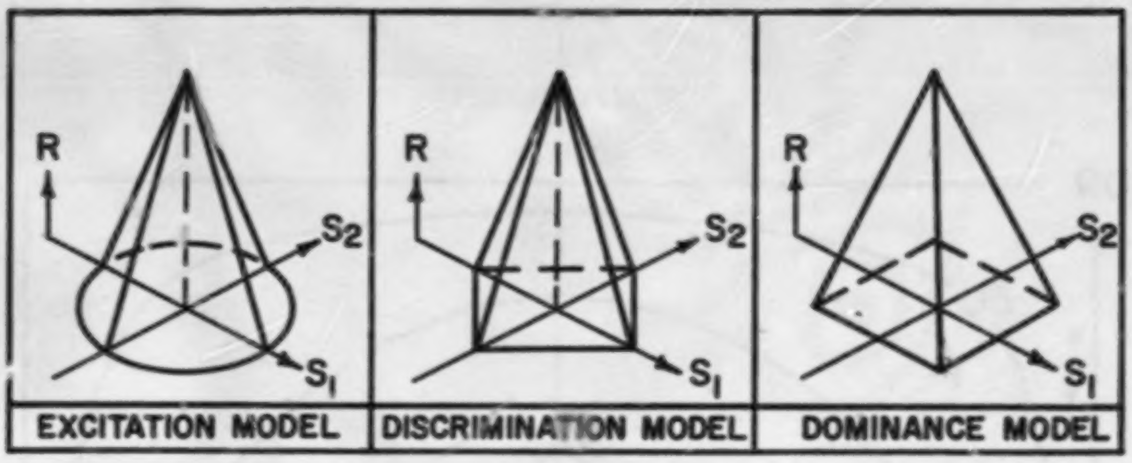

R
S2
S1
EXCITATION MODEL
R
S2
S1
DISCRIMINATION MODEL
R
S2
S1
DOMINANCE MODEL

Fig. 1

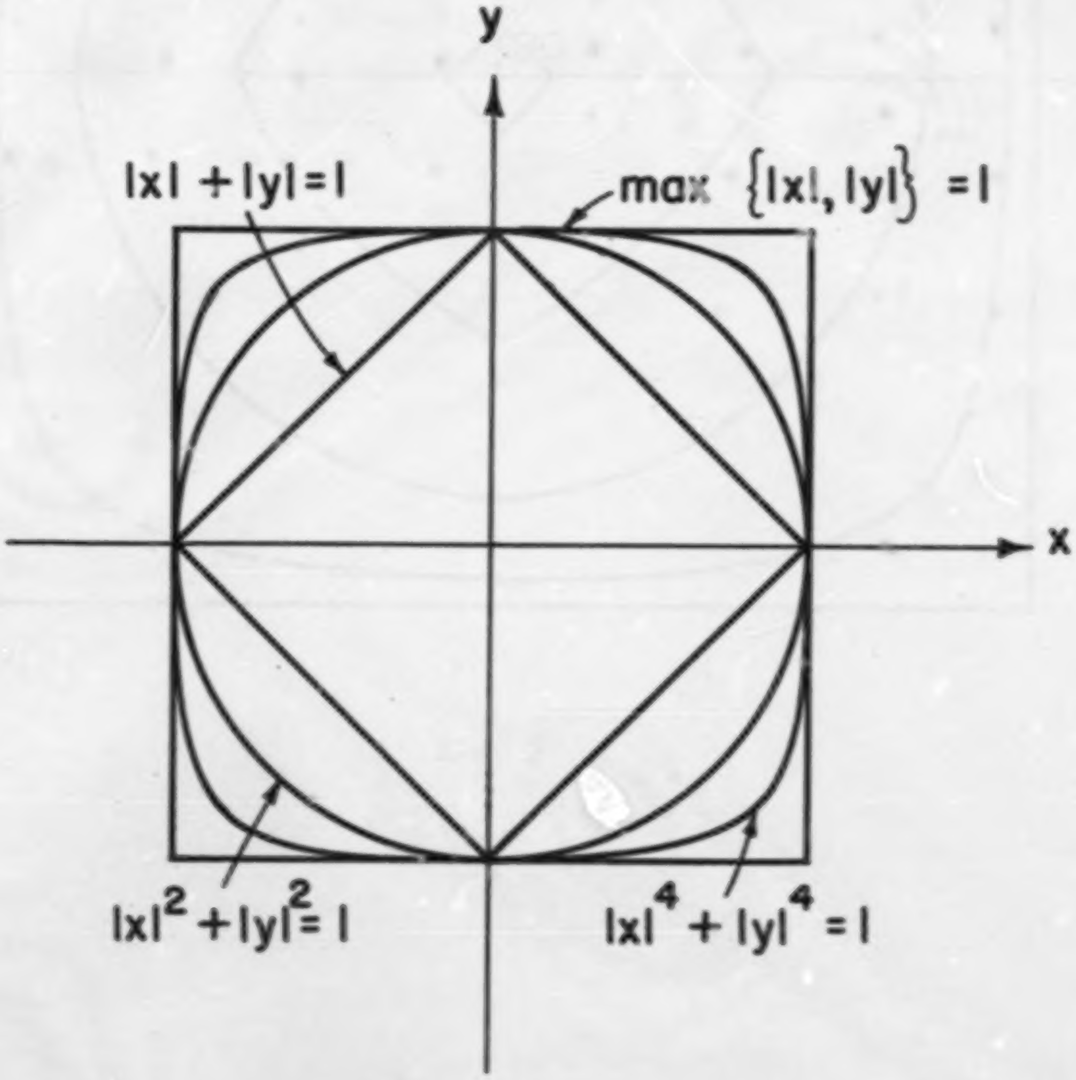

y
x
|x| + |y| = 1
max {|x|, |y|} = 1
|x|² + |y|² = 1
|x|⁴ + |y|⁴ = 1

Fig. 2

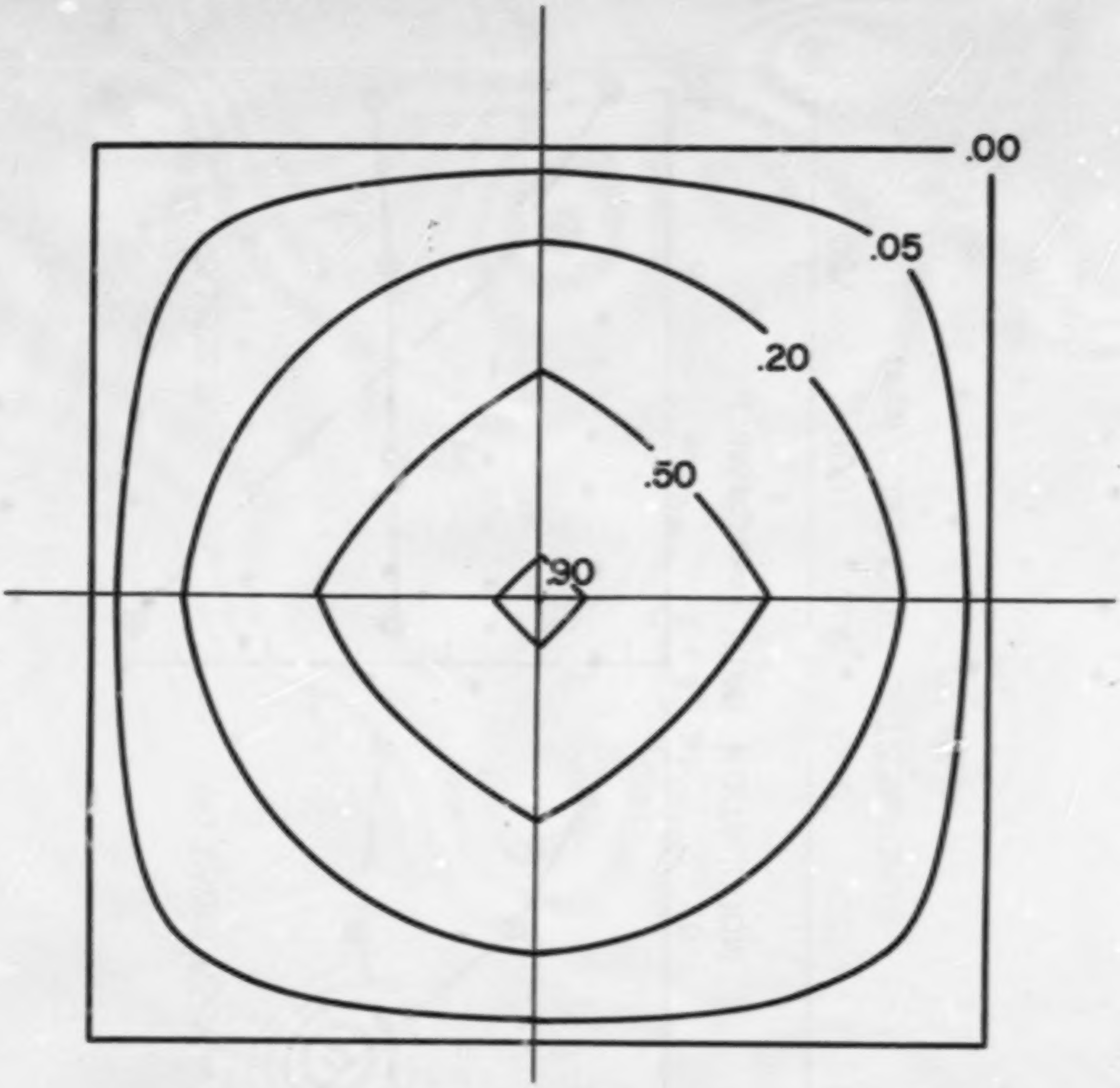

.00
.05
.20
.50
.90

Fig. 3

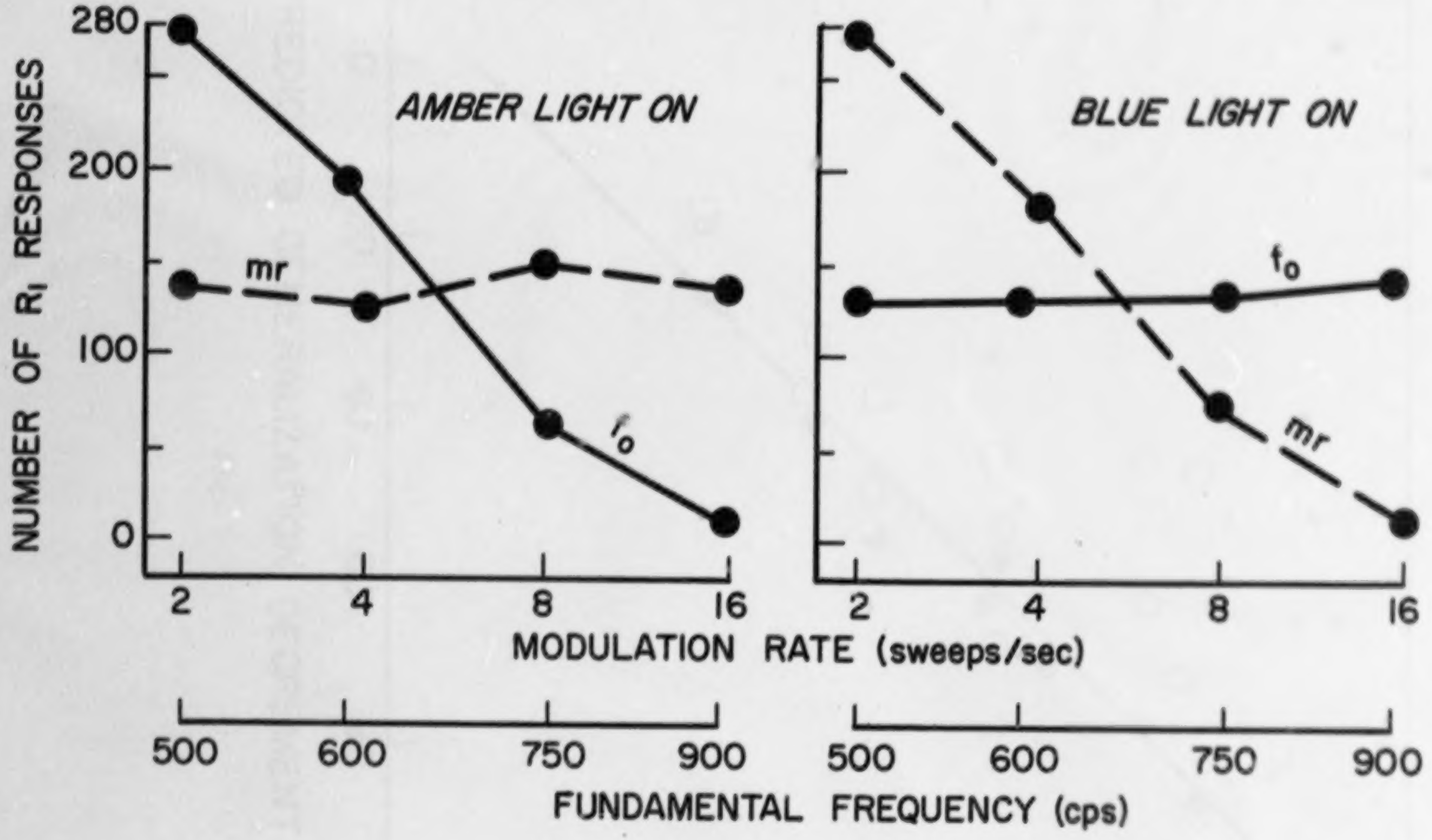

NUMBER OF R_1 RESPONSES
280
200
100
0
AMBER LIGHT ON
mr
f_o
BLUE LIGHT ON
f_o
mr
2
4
8
16
2
4
8
16
MODULATION RATE (sweeps/sec)
500
600
750
900
500
600
750
900
FUNDAMENTAL FREQUENCY (cps)

Fig. 4

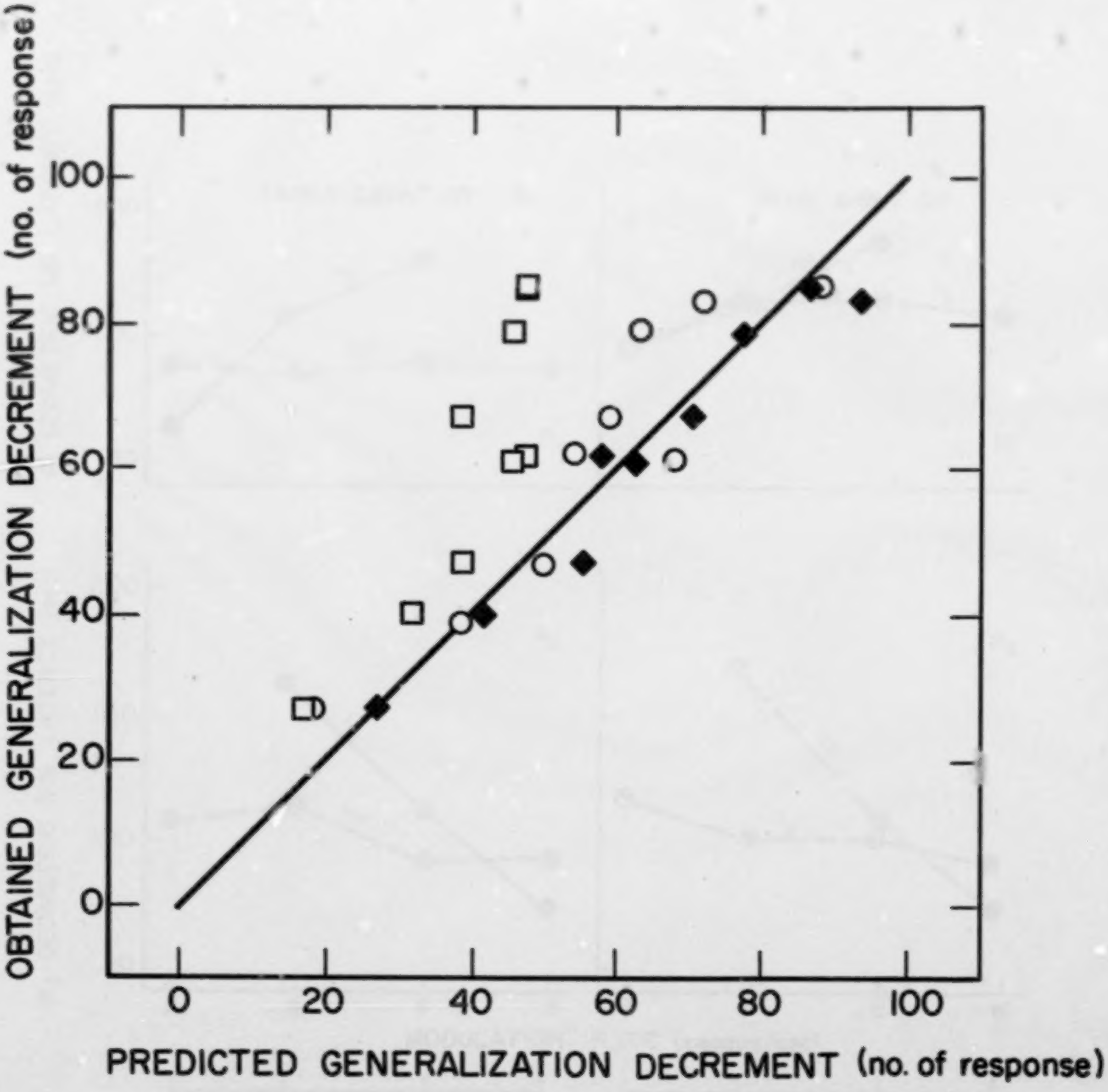
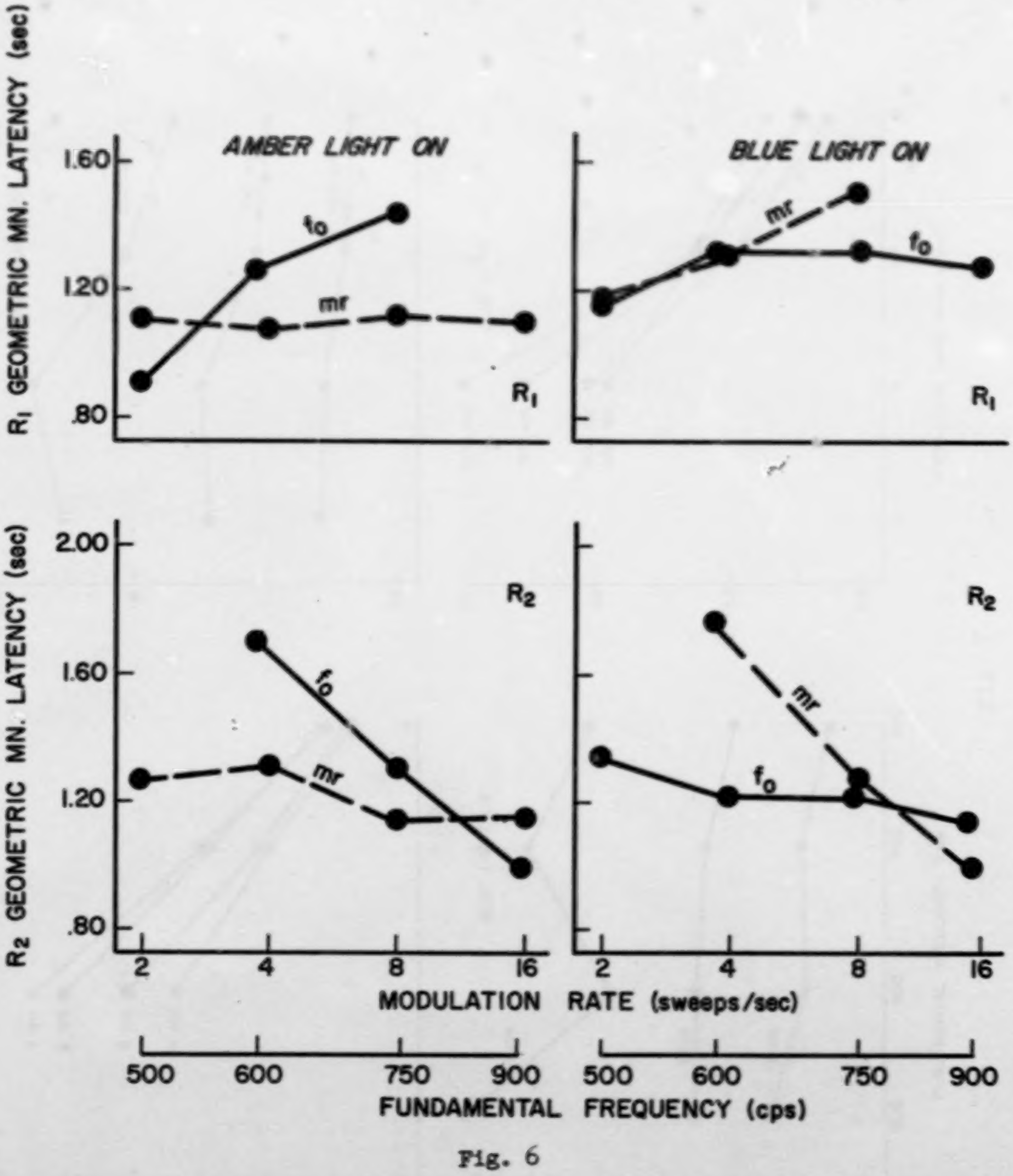

OBTAINED GENERALIZATION DECREMENT (no. of response)
100
80
60
40
20
0
0
20
40
60
80
100
PREDICTED GENERALIZATION DECREMENT (no. of response)

Fig. 5



AMBER LIGHT ON
BLUE LIGHT ON
R₁ GEOMETRIC MN. LATENCY (sec)
1.60
1.20
.80
f₀
mr
R₁
R₂ GEOMETRIC MN. LATENCY (sec)
2.00
1.60
1.20
.80
R₂
2
4
8
16
MODULATION RATE (sweeps/sec)
500
600
750
900
FUNDAMENTAL FREQUENCY (cps)

Fig. 6

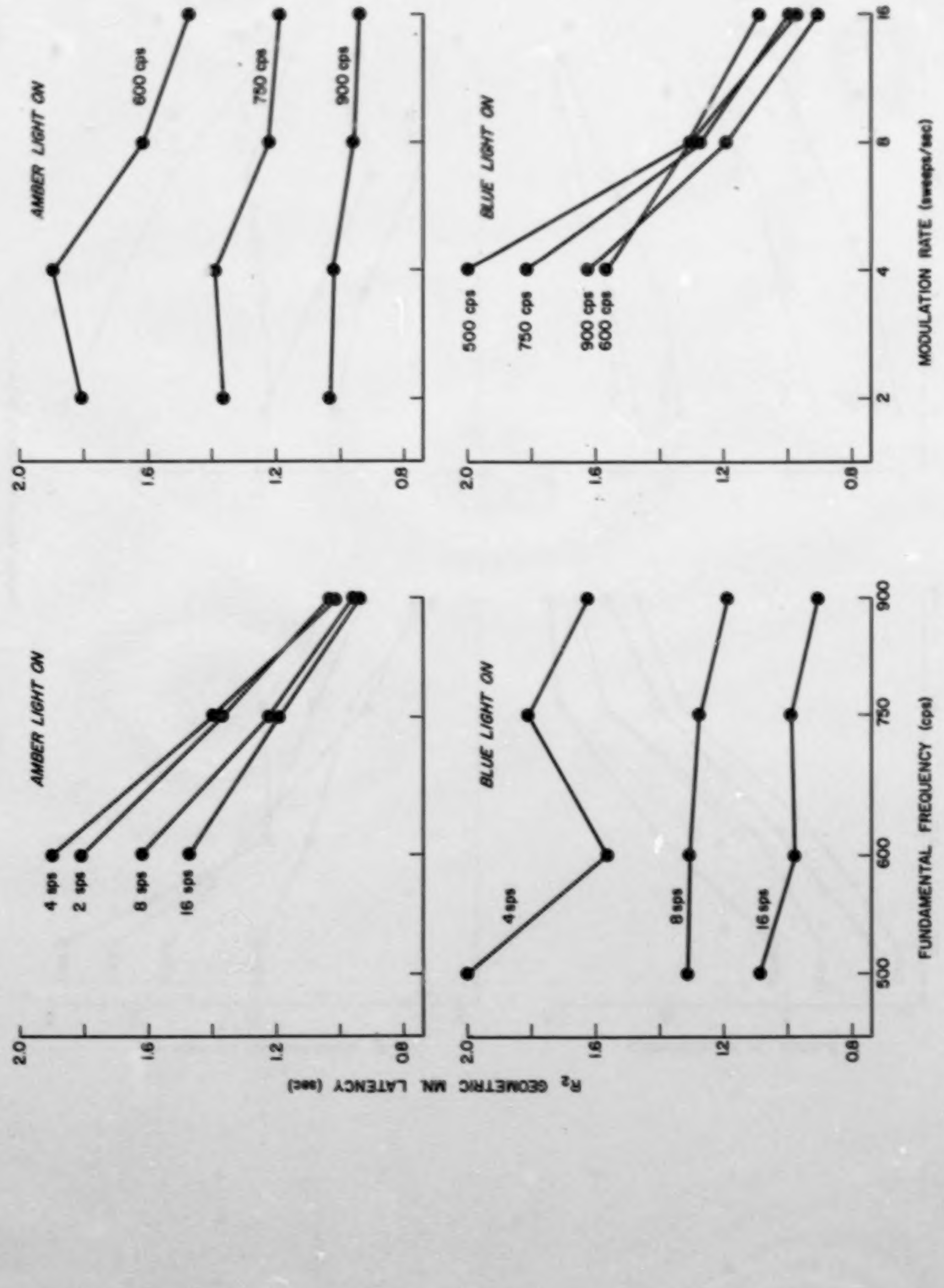
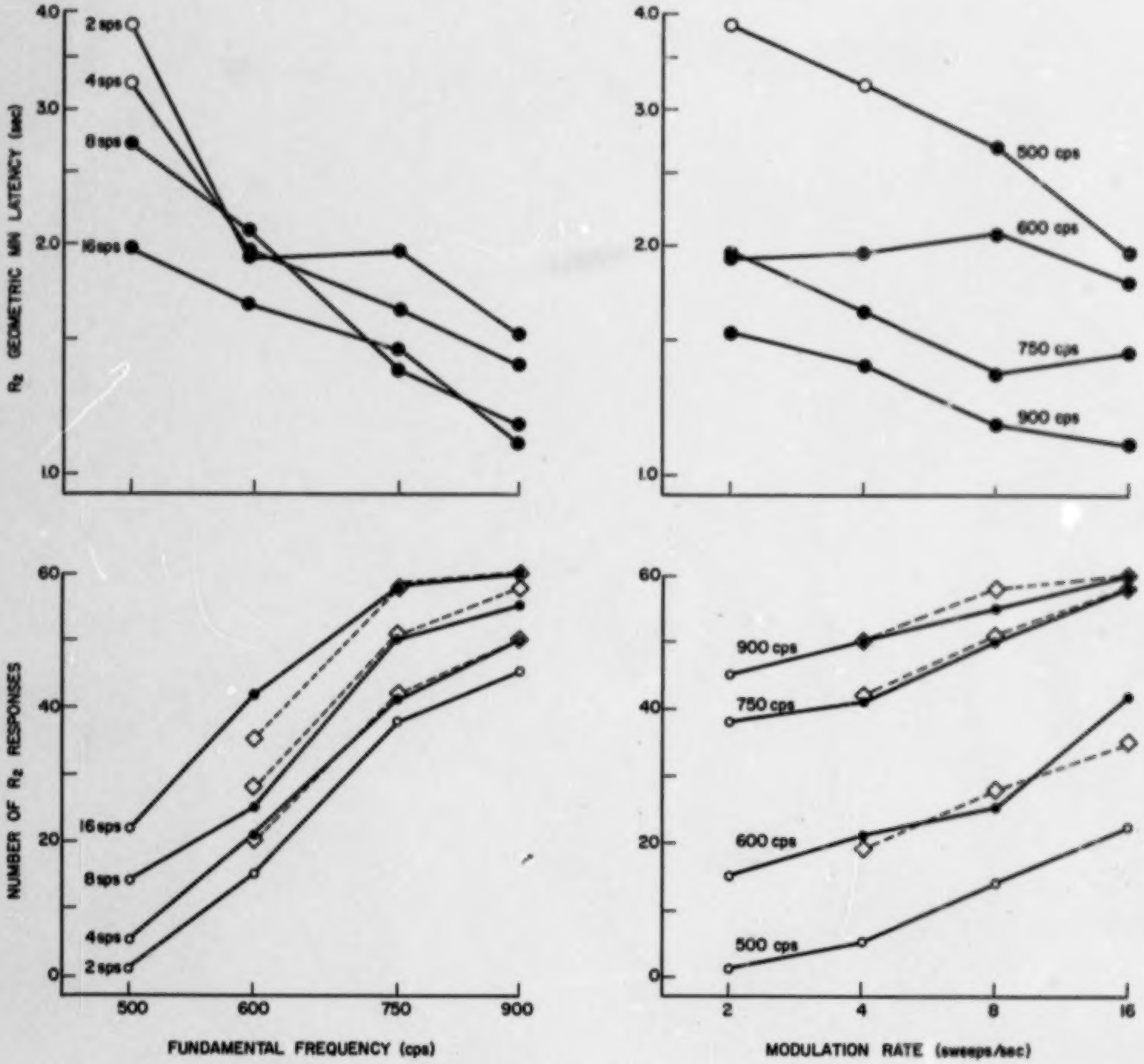

AMBER LIGHT ON
4 sps
2 sps
8 sps
16 sps
BLUE LIGHT ON
4 sps
8 sps
16 sps
FUNDAMENTAL FREQUENCY (cps)
500
600
750
900
R₂ GEOMETRIC MN. LATENCY (sec)
2.0
1.6
1.2
0.8
AMBER LIGHT ON
600 cps
750 cps
900 cps
BLUE LIGHT ON
500 cps
750 cps
900 cps
600 cps
MODULATION RATE (sweeps/sec)
2
4
8
16

Fig. 7



AMBER AND BLUE LIGHTS BOTH ON

Fig. 8

THE EFFECTS OF RESPONSE-DEPENDENT AND INDEPENDENT REINFORCEMENT IN EXTENDING STIMULUS CONTROL

Harlan Lane and James L. Kopp
The University of Michigan

The Effects of Response-Dependent and Independent Reinforcement in Extending Stimulus Control

Harlan Lane and James L. Kopp
Behavior Analysis Laboratory
The University of Michigan

Covert responding is generally considered to differ in magnitude but not in kind from overt behavior. Electromyography permits measurement of covert skeletal responses of exceptionally small magnitude. The limited evidence to date suggests that covert responses, like their more molar counterparts, may be controlled by discriminative stimuli and by reinforcement (Hefferline, 1959).

The present study undertakes a description of reinforcement control of covert vocal responding in human adults. We attempt to circumvent the formidable problems of measurement by presenting reinforcement under conditions in which the covert response is likely to occur but is not known to occur. We then test for the effects of this "conditioning" on the overt counterparts of the covert responses.

The likelihood of the covert response is arranged through discrimination training of its overt counterpart; the effect of response-independent reinforcement is assessed by subsequently measuring the generalization gradient for the overt responses.

The discrimination training--generalization testing paradigm is adapted from that employed by several recent studies (LaBerge, 1961; Sherman & Pierrel, 1961). Two stimuli are selected from a physical continuum and presented during conditioning; one of the stimuli (S^D) is correlated with reinforcement of an operant. In generalization testing other stimuli from this continuum are presented recurrently and the frequency of responding is noted in the absence of reinforcement. The gradient of generalization, relating response frequency to stimulus magnitude, is characterized by plateaus in the regions of S^D and S^{Δ} separated by a relatively steep gradient of response probability over intermediate stimulus values.

A subsequent study by Cross and Lane (1962) elaborated this paradigm. These authors employed two responses (R_1, R_2) conditioned to two discriminative stimuli ($S^D{}_1$, $S^D{}_2$). R_1 is reinforced in the presence of $S^D{}_1$ and not in the presence of $S^D{}_2$, and the complementary contingencies obtain for R_2. This technique yields two reciprocal generalization gradients of the same form as in the single-response procedure described above. Cross and Lane also found a simple relation between these reciprocal gradients of response frequency and the latnecy of these responses: latency is maximum when response probabilities are equal and decreases as the difference between the probability of the two responses increases. Kellog (1931) has reported comparable data on response latency.

The disjunctive response--dual stimulus procedure was adapted to investigate covert behavior in the following way. After initial discrimination training, the experimental group of subjects was given additional reinforcement during the presentation of stimuli which were intermediate on the stimulus continuum between S_1 and S_2. An overt verbal response was not required. Reinforcement was given following every stimulus presentation without regard to the _S_'s behavior. During this period of additional conditioning to stimuli adjacent to the original S^D's, any covert responses emitted should be strengthened by their reinforcing consequences. The effects of this conditioning may then be compared to (a) effects of reinforcing overt behavior during this period and (b) effects of omitting this interpolated period, by a comparison of the generalization gradients for the three groups.

In general, we would expect that extending discrimination training to stimuli adjacent to the original S^D's would increase the slope of the response frequency gradients and of the response latency gradient. This hypothetical outcome is shown in Fig. 1.

Method

Subjects and Stimuli

Thirty-six volunteer undergraduates, assigned to groups of 12 randomly, served individually in sessions lasting approximately 45 min. The subject was seated in an audiometric room in front of a counter, signal light, and microphone. Auditory stimuli were presented monaurally through a binaural headset with calibrated earphones (PDR-8). The stimuli were 500-cps tones, 1.2 secs in duration, recorded on magnetic tape. In order to reduce audible print-through and noise during playback of the recordings, a narrow band-pass filter (Krohn-Hite 310 AB) was interposed between the tape recorder output (Uher 3-A) and the headphones.

The taped signal also tripped a voice-operated relay (Miratel) which started an electronic counter (Hewlett-Packard 522B). The subject's response to the stimulus operated another voice-relay which, in turn, stopped the counter. The start-stop interval was read in milliseconds from the counter and taken as latency. (All control apparatus was located outside the experimental booth.)

The vocal responses employed were the phonetic clusters /ka/ and /ti/. Topographically discrete responses were selected to minimize response generalization.

Procedure

Phase 1. Training (all groups). After S was seated in the experimental booth the following instructions were given through an intercom:

"In this experiment you can earn money by simply saying /ka/ or /ti/ at appropriate times. We can't tell you now when or how these responses should be used. That is for you to learn. All you have to do is wear the headphones, and watch the display unit in front of you. You will hear various sounds. Each time you respond appropriately the green light will flash and five points will be added to your score on the counter. You will want to get as high a score as possible because the amount we pay you at the completion of the experiment will be determined by your final score." (Questions were answered by repeating the instructions.)

One hundred 500-cps tones were presented to S; the tones were randomized in blocks of 20, half at 56 db and half at 74 db (SPL). The 56 db tone was the discriminative stimulus (S^D_1) for a /ti/ response (R_1) and the 74 db tone was the discriminative stimulus (S^D_2) for a /ka/ response (R_2). S^D_1 was the stimulus consistently correlated with non-reinforcement (S^{Δ}_2) for R_2, and S^D_2 was the stimulus consistently correlated with non-reinforcement (S^{Δ}_1) for R_1. During the first 20 trials, reinforcement was presented if a single /ti/ response followed S^D_1 or a single /ka/ response followed S^D_2, within the allowed time interval (5.5 secs). A partial reinforcement schedule was employed after the 20th trial, in order to minimize extinction in later testing. Each response was reinforced on one-third of its correct occurrences; the schedule was adjusted, however, to insure that both responses would be reinforced an equal number of times.

Phase 2. Interpolated phase. (Experimental Group I). The subject was instructed as follows:

"For this next part of the experiment you are asked to remain as you are. You will again hear various sounds. Remain attentive and watch the display unit but from now on do not say anything aloud into the microphone."

One hundred 500-cps tones were presented to S, 25 each at 62, 65, 68, and 71 db (SPL), randomized in blocks of 20. These intensities were intermediate between the 56 db and 74 db stimuli of Phase 1.

Reinforcement was given approximately 2 secs after the onset of each stimulus and in the absence of overt responding by S. The two-sec delay of reinforcement was close to the maximum average latency for vocal responding of Ss in the Cross and Lane study.

Phase 2. Interpolated phase. (Experimental Group II). The subjects' instructions were:

"For this next part of the experiment you are asked to remain as you are. You will again hear various sounds. Remain attentive, watch the display unit and respond as you think appropriate."

The same stimuli were administered as to Experimental Group I during Phase 2. However, reinforcement was given either after each vocal response or 4 sec after the onset of the stimulus, whichever occurred first. In this condition it was necessary to delay the reinforcement 4 sec in order to allow S ample time to make a response.

Phase 3. Testing (all groups). In this final phase, Ss were told they could continue to earn points for /ka/ and /ti/ responses but that the points would accumulate on a counter outside the chamber, their own unit becoming inoperative. One hundred and ten stimuli were presented in random order at 11 intensity levels (3 db steps from 50 through 80 db, SPL).

Results

Examination of stimulus generalization gradients for individuals suggested that the behavior of some of the Ss was not under discriminative control. In order to facilitate comparison of the generalization gradients produced by the experimental treatments, the data from those Ss who gave less than 60% correct responses to the last block of 20 stimuli in the training phase ("sub-criterion group") are treated separately.

Figure 2 summarizes response probabilities and Fig. 3 response latencies for the three groups and the sub-criterion group during generalization testing. All groups yielded data of essentially the same form as that found by Cross and Lane. In particular, the maximum number of R_1 and R_2 responses were at the extreme low and high ends of the continuum, respectively, and not at the two S^D intensities (Fig. 2). The gradients were asymmetric in the direction of the more intense stimuli. For the groups reaching criterion in Phase 1, the average latencies for all responses are minimal when the difference in response probabilities is greatest (Fig. 3). The latency function reaches a maximum when the probabilities of the two responses approach equality.

As to differences due to extended reinforcement for the experimental groups, it can be seen from Fig. 2 that the group reinforced during the intermediate phase with no verbal responding (E-I) did not differ from the control group (no interpolated phase) in the form and position of generalization gradients. The group which was reinforced while responding overtly during phase 2 (E-II), yielded gradients of essentially the same form as the control and E-I groups but the point of 50% response probability was shifted about 3 db toward the more intense end of the stimulus continuum.

Generalization gradients for the sub-criterion group were less regular. Although the form of these gradients indicates some degree of differential stimulus control of responding, the probability of the dominant response never went beyond .75.

The average latencies of responding for each of the three groups were essentially the same in minimum and peak values. All latencies tended to peak at about the intensity where the reciprocal generalization gradients intersected. The latencies for the sub-criterion group were much longer on the average and showed only a slight relationship to response probabilities at maxima and minima of the latency functions.

Separate latency functions for each of the two responses are shown in Fig. 4. Since the average latency functions for all groups (except the sub-criterion group) were essentially the same, the data shown in Fig. 4 for the control group are representative. Latencies accompanying the stochastically dominant response (the response with the highest probability of occurrence at a given stimulus intensity) were consistently shorter than the latencies associated with the non-dominant response. Both latency functions increase to a maximum at a point displaced 12 db from their respective S^D intensities toward the center of the continuum, then decrease systematically. The same trend is noticed in plots of separate latency functions for individual subjects (Fig. 5).

Discussion

The results of this experiment add weight to the generalization that multiple response discrimination training effectively divides the stimulus continuum into sharply defined, response-specific categories or classes. As in the study by Cross and Lane, the mean latency for all responses at a given stimulus value varied approximately linearly with the variance of the obtained distribution of response probabilities; also, the latency of the stochastically dominant response was consistently shorter than that associated with the non-dominant response.

Since the generalization gradients obtained for the group (E-I) of Ss who received an interpolated phase of regular reinforcement with no verbal response required, seemed remarkably congruent with these data from the control group, it is concluded that, under the conditions of this experiment, reinforcement that is not contingent on an overt response has little effect on the control of discriminative behavior.

The generalization gradients for the group (E-II) of Ss, who responded overtly and received 100% reinforcement in phase 2, reveals a shift toward the more intense end of the stimulus continuum. Analysis of individual data showed that, in the majority of cases, S emitted one or the other of the two responses consistently in the face of the uniform (100%) reinforcement. In these cases R_1 was generally the stereotyped response. The stereotypy of responding under these conditions appears similar to the "superstitious behavior" found in infra-human experiments employing a reinforcing stimulus which is not response-contingent (Skinner, 1948). The preference for R_1 may be explained by the fact that the first stimulus in phase 2 was only 6 db from the S^D for R_1 in the training phase and was 15 db removed from the S^D for R_2. This was true for all Ss. Thus, stimulus generalization initially gave R_1 its dominance, which was then retained in the form of "superstitious" responding due to 100% response-independent reinforcement. The effect of conditioning this stereotyped behavior is reflected in the generalization gradients for E-II.

V

The slight shift toward the more intense end of the continuum of the point of intersection for the gradients of the control group and E-I are in accord with the findings of Cross and Lane. This shift is probably not to be explained in terms of response bias: analysis of the responses of these groups during phase I showed no consistent response preferences either by the individual Ss or by each group as a whole. The asymmetry is comparable to that obtained in investigations employing single response procedures and has been discussed by Hull (1949) in terms of an interaction between "stimulus-intensity dynamism (V)" and "stimulus-intensity generalization (SER)."

Summary

Auditory generalization gradients of response probability and latency were obtained from human Ss under three conditions: (a) discrimination training was given to 500 cps tones at 56 and 74 db, and followed by generalization testing over eleven different intensity levels at 3 db intervals from 50 to 80 db (SPL); (b) an additional phase was interpolated between discrimination training and generalization testing, in which stimuli adjacent to the training stimuli were presented (62, 65, 68, and 71 db). All stimuli were followed by reinforcement and there was no overt responding during the interpolated phase; (c) the treatment was identical with that of (b), except that S gave an overt response during the interpolated phase.

The shape and position of both response-probability gradients and latency functions substantiated the findings of an earlier study (Cross and Lane, 1962) on the discriminative control of concurrent responses.

Differences in discrimination training failed to differentiate in generalization testing the group which received only the initial training phase and was then tested (control group) from the group which was presented with the additional stimuli of the interpolated phase and reinforced after each stimulus, independent of responding. The third group of Ss, who were given the additional stimuli of the interpolated phase and regularly reinforced while responding overtly, yielded similar gradients with a marked shift toward the more intense end of the continuum. The cause of the shift was traced to a response bias developed superstitiously during the response-independent, regular reinforcement condition.

References

Cross, D. V. & Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization. *J. exp. anal. Behav.*, 1962, 5, 487-496.

Hefferline, R. F., Keenan, B., & Harford, R. A. Escape and avoidance conditioning in human subjects without their observation of the response. *Science*, 1959, 130, 1338-1339.

Hull, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, 56, 67-76.

Kellog, W. N. The time of judgment in psychometric measures. *Amer. J. Psychol.*, 1931, 43, 65-86.

LaBerge, D. Generalization gradients in a discriminative situation. *J. exp. Psychol.*, 1961, 62, 88-94.

Sherman, J. C. & Pierrel, R. Generalization of auditory intensity as a function of amount of discrimination training. *J. exp. anal. Behav.*, 1961, 4, 237-241.

Skinner, B. F. 'Superstition' in the pigeon. *J. exp. Psychol.*, 1948, 38, 168-172.

Figure Captions

Fig. 1. Hypothetical response-frequency and latency gradients for a control and an experimental group, when the latter receives additional discrimination training to intermediate stimuli.

Fig. 2. Conditional probabilities of R_1 and R_2 at each stimulus intensity for each group. Probabilities were estimated from the total number of /ka/ responses (triangles) and /ti/ responses (circles) emitted in 10 presentations of each stimulus intensity.

Fig. 3. Average latency of responding (R_1 and R_2 combined) at each stimulus intensity. The total latency at each stimulus is the unweighted mean of the average latency of responding for Ss in each of the four groups.

Fig. 4. Average latency of responding (R_1 and R_2 computed separately) at each stimulus intensity for the control group only. Each point represents the average latency of the indicated number of R_1 (triangles) and R_2 (circles) responses.

Fig. 5. The average latency of responding (R_1 and R_2 computed separately) at each stimulus intensity for individual Ss in the control group.

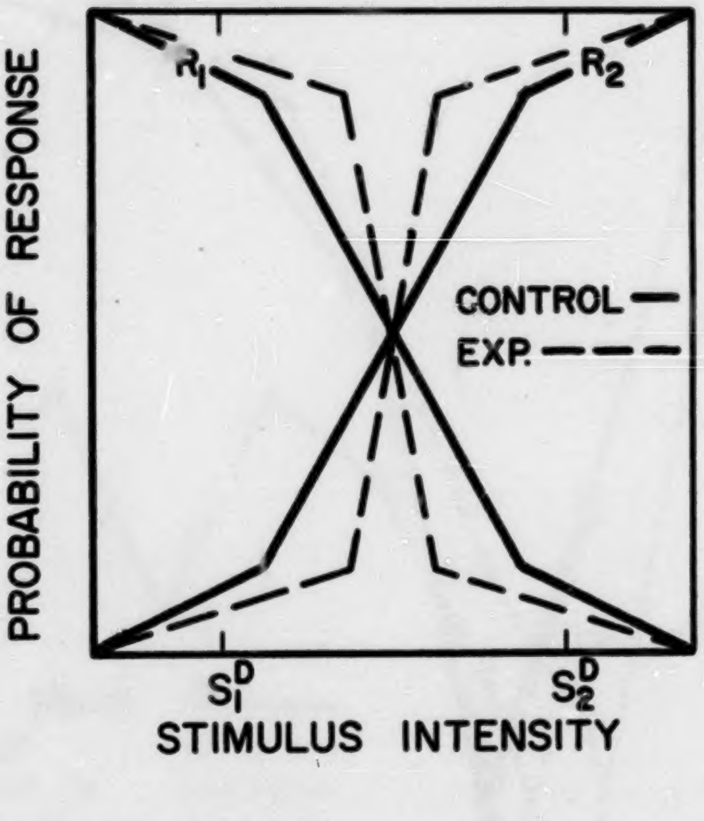

PROBABILITY OF RESPONSE
R_1
R_2
CONTROL
EXP.
S_1^D
S_2^D
STIMULUS INTENSITY
MEAN LATENCY
CONTROL
EXP.
S_1^D
S_2^D
STIMULUS INTENSITY

Fig. 1

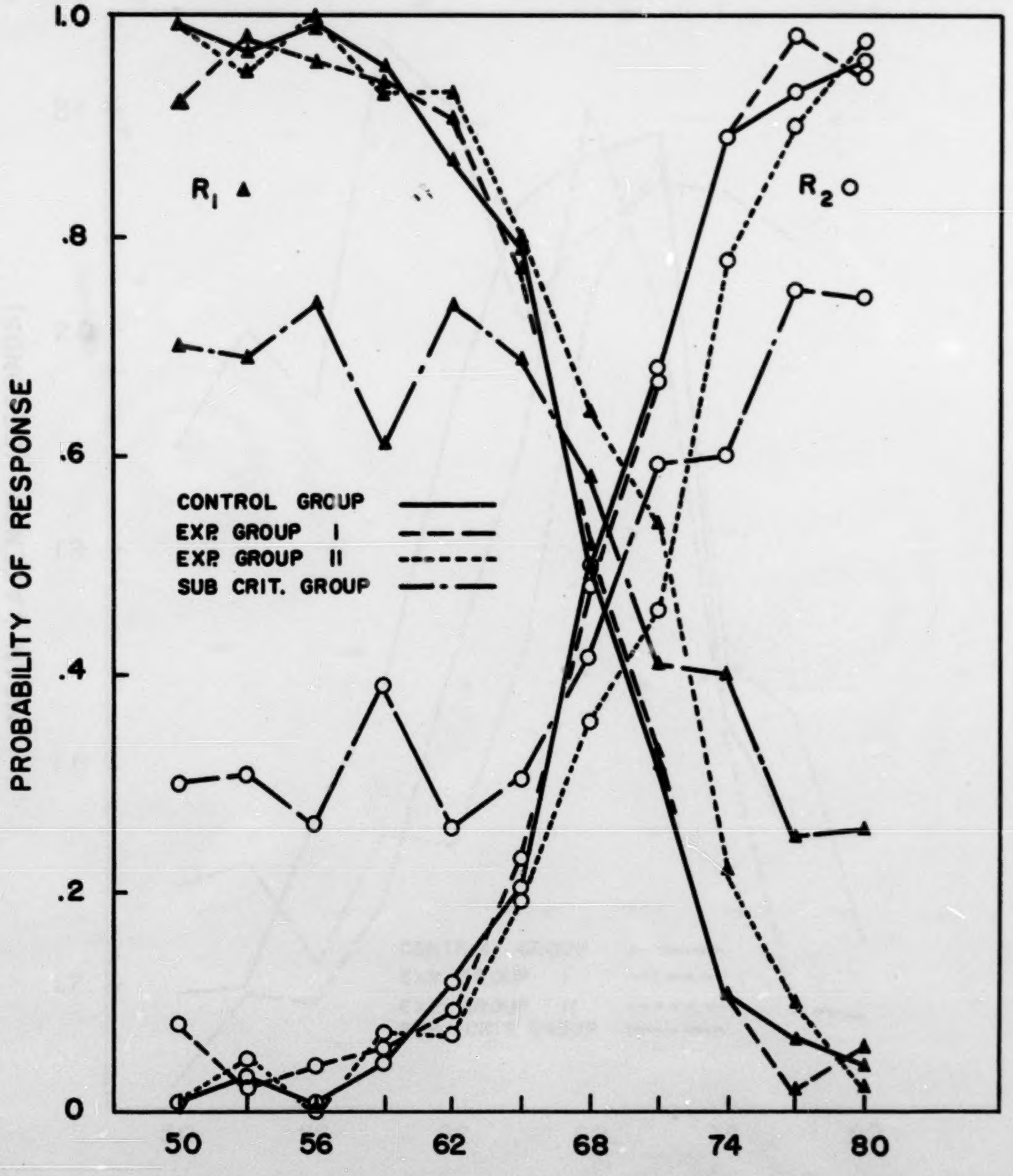

1.0
.8
.6
.4
.2
0
PROBABILITY OF RESPONSE
R₁ ▲
R₂ ○
CONTROL GROUP
EXP. GROUP I
EXP. GROUP II
SUB CRIT. GROUP
50
56
62
68
74
80
STIMULUS INTENSITY (db RE: .0002 DYNES/CM²

Fig. 2

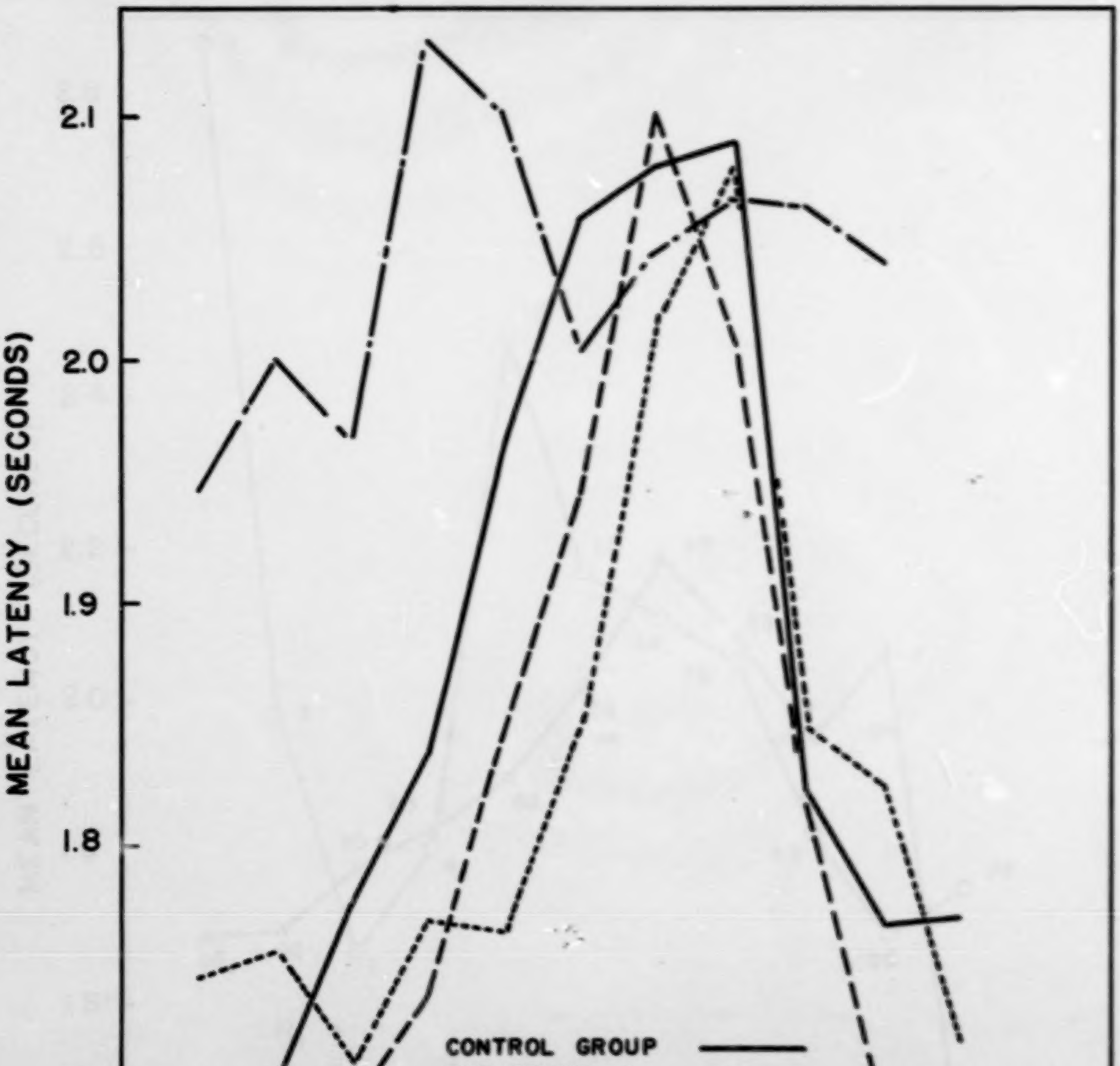

MEAN LATENCY (SECONDS)
2.1
2.0
1.9
1.8
1.7
50
56
62
68
74
80
CONTROL GROUP
EXP. GROUP I
EXP. GROUP II
SUB CRIT. GROUP
STIMULUS INTENSITY (db RE: .0002 DYNES/CM²)

Fig. 3

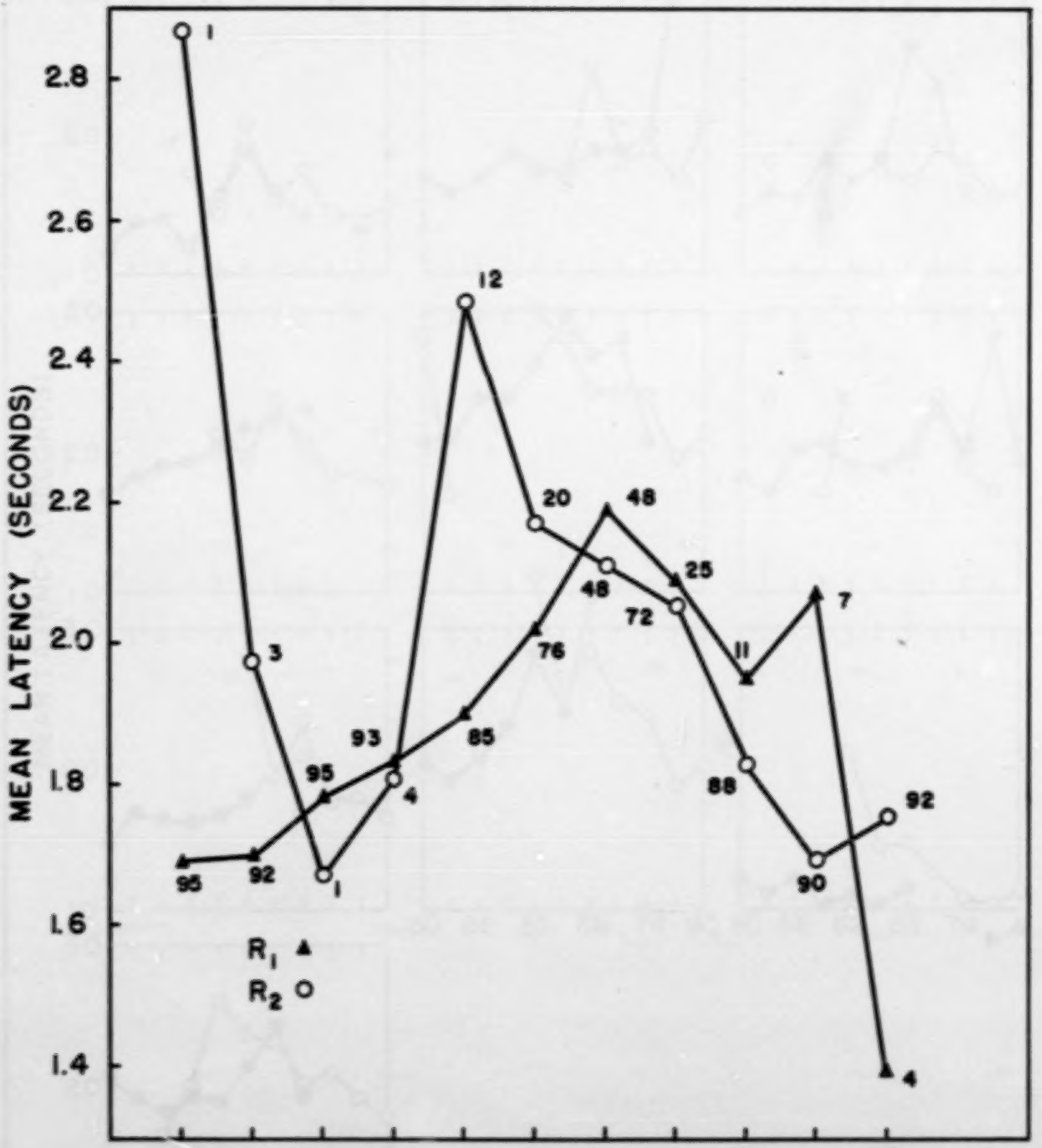

MEAN LATENCY (SECONDS)
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
50
56
62
68
74
80
STIMULUS INTENSITY (db RE: .0002 DYNES/CM^2)
1
3
1
4
12
20
48
72
88
90
92
95
92
95
93
85
76
48
25
11
7
4
R_1 ▲
R_2 ○

Fig. 4

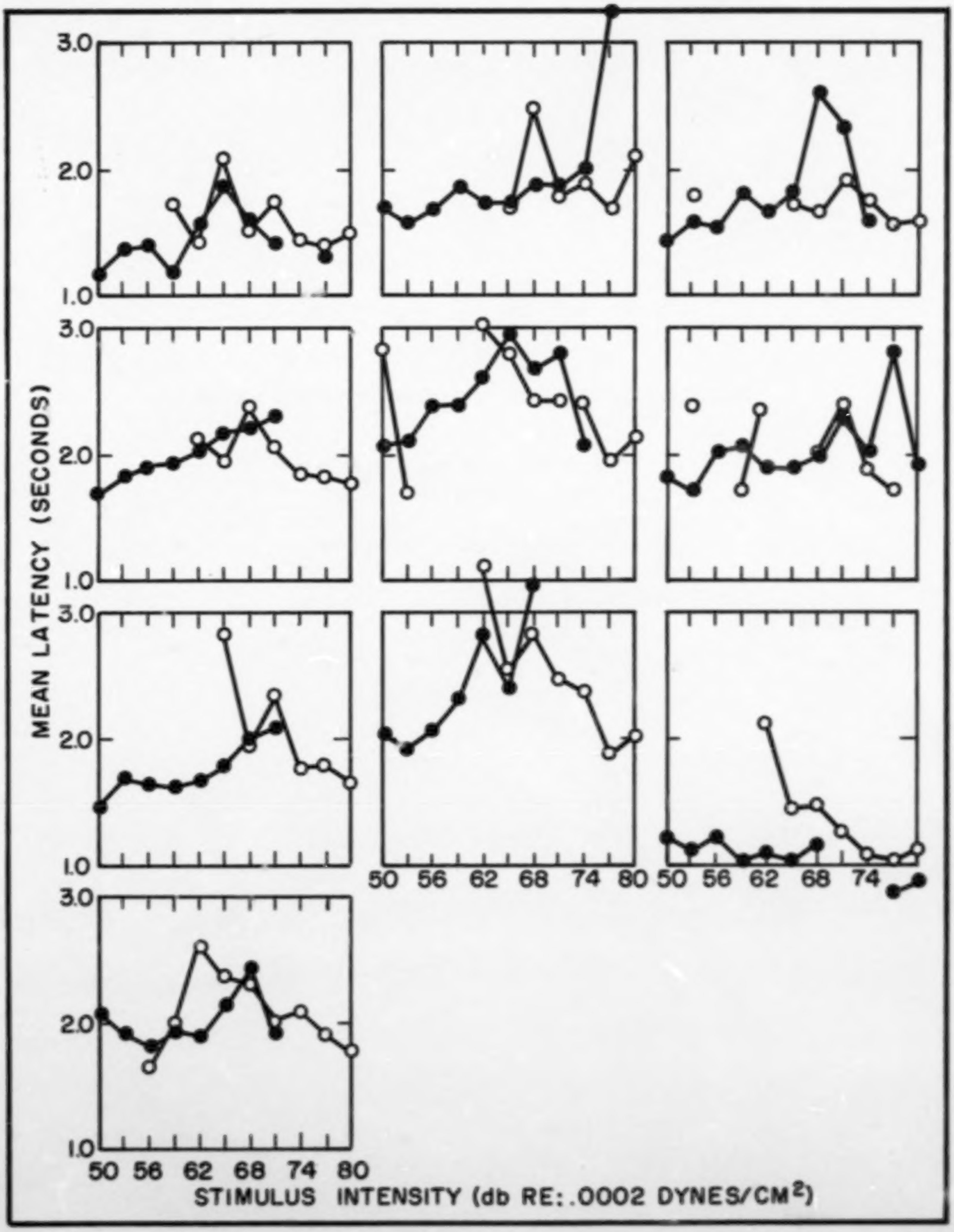

3.0
2.0
1.0
MEAN LATENCY (SECONDS)
50
56
62
68
74
80
STIMULUS INTENSITY (db RE: .0002 DYNES/CM²)

Fig. 5

SOME DISCRIMINATIVE PROPERTIES OF SYNTACTIC STRUCTURES

Harlan Lane and Bruce Schneider
The University of Michigan

Some Discriminative Properties of Syntactic Structures

Harlan Lane and Bruce Schneider
Behavior Analysis Laboratory
University of Michigan

In a recent article discussing some psychological studies of grammar, Miller (1962) reports an unpublished experiment by Jacques Mehler in which subjects are given a list of sentences, varying both in content and structure, to read and then recall. The errors in recall after the first few trials are rarely due to omissions and semantic confusions; the bulk of the errors occur because S alters the syntactic form of the sentence.

Miller suggests that the process of recall begins with a recoding of the sentence into the kernel plus "some kind of implicit code" of the transformation that is necessary to reconstruct the sentence from the kernel on demand. A maximum of three transformations plus the kernel were needed to reconstruct certain of the sentences that Mehler presented. Thus, Miller proposes that "the sentences were recoded and that each of the four components of the kernel-plus-code was remembered correctly or incorrectly independently of the others." This hypothesis leads to a theoretical distribution of syntactic errors which accords well with that obtained.

A critical step in the recoding of sentences into kernel plus transformation is the discrimination of the syntactic structures involved. One way of exploring the discrimination of syntactic sturctures is to give a listener some task which requires discriminations among sentences generated by different transformations of the same kernel. In this experimental setting, we may obtain information on the discriminability of syntactic structures, that is, the degree to which they do not overlap in controlling differential responding. This is the first objective of the present study.

The degree of overlap or "confusion" of syntactic structures may be expected to depend greatly on their relative frequencies within a corpus. If one structure predominates, it may preempt discriminative control, and errors in responding will be associated primarily with the less frequent structures. When the several structures approach equal frequency there are three possible outcomes. First, the S may respond appropriately to each. In Mehler's experiment involving recall, Miller reports that "the Ss quickly get the impression that about half the sentences are negative, half are passives, half are questions; in recall, therefore, they try a little probability matching." Second, a particular structure may appropriate control. Third, discriminative control may break down altogether and the S respond randomly with respect to the stimulus array (the corpus).

The second objective of the present study, therefore, is to examine the effects of the relative frequencies of the syntactic structures in a corpus on the discriminative control exerted by those structures.

Method

Preparation of the stimulus array

Negative, passive, and querry transformations were applied to 36 kernel sentences, yielding 144 sentences sampling four syntactic structures. As an example:

The people support Kennedy is a declarative sentence (D).

The people do not support Kennedy is a negative sentence (N).

Kennedy is supported by the people is a passive sentence (P).

Do the people support Kennedy? is a querry (Q).

The 144 sentences were assigned to each of ten 4 x 4 matrices (Table 1). The rows of each matrix represent four speakers, the columns represent the four syntactic structures, and the cells give the number of sentences of each form uttered by a single speaker. Each speaker read every one of the 36 kernel sentences in one of its four forms. In matrix 1, speaker 1 read all D-sentences, speaker 2 all N-sentences, etc. In matrix 2, speaker 5 read 33 D-sentences and one each of the other forms; similarly speakers 6 through 8 read the corresponding dominant structure (N, P, Q, respectively) 33 times and the non-dominant structures once each. In successive matrices in the series of 10, the relative frequency of the dominant structure in each row decreased, while the non-dominant structures increased in frequency but remained equal to each other. The uncertainty of the stimulus array ranged from 2 to 4 bits.

The 36 sentences comprising the corpus of each speaker were arranged in random order both with respect to the distribution of kernels and the distribution of syntactic structures determined by the matrix. Forty male speakers (four speakers for each of the ten matrices) read the sentences with conversational intonation while seated in front of a microphone in a radio studio. Each corpus was recorded on a loop of magnetic tape, labelled, and catridged separately.

Subjects and procedure

Six groups of five Ss each served in separate sessions lasting five hours. A 15 minute hiatus was provided after 2 and 4 hours; Ss were not permitted to discuss the experiment.

The four corpora comprising a single matrix and the 10 matrices were presented in random order to the groups of Ss. The following instructions were read:

"This is an experiment in language perception. What we would like you to do is to listen carefully to passages from four different speakers. The first speaker you hear will be called speaker 1. The second will be called speaker 2, etc. After you have heard the four speakers each of you will be presented with a deck of index cards. On each card will be typed a sentence. What we would like you to do is to assign these sentences to one of the four speakers; speakers 1 through 4. You are to assign a sentence to a speaker if you think that he was more likely to have said it than the other speakers. The index cards will be turned face down on the table. You are to turn them over one at a time and assign them to one of four piles, the piles belonging to the respective speakers. Arrange the piles so that speaker 1 is on the left and speaker 4 on the right.

Remember, in order to assign sentences correctly, it is important that you listen carefully to each speaker and that you be sure that the sentences you assign, say, to the pile labeled "speaker one" (far left) are those which you believe are more likely to have been said by the first speaker. In the same manner be sure that the sentences that you assign to the pile labeled "speaker two" are those more likely to have been said by the second speaker. This of course goes for speakers 3 and 4.

Are there any questions? Then we will begin."

After listening to a set of four speakers, comprising a matrix, S was given a deck of index cards containing 144 different sentences, representing 36 occurrences of each syntactic structure in random order. Ten such decks were prepared from 1440 simple sentences; a new deck was selected at random and presented to S after each matrix. The Ss were visually isolated from each other and were not constrained in the time to complete the sorting task. The cards were collected so as to preserve the assignment of sentences to speakers.

Ten response (R) matrices were prepared later for each S. The matrices showed the frequency with which each of the four sentence structures was assigned to each of the four speakers in the corresponding stimulus (S) matrix. Comparison of the S- and R- matrices revealed the degree of discriminative control, while comparison of the cells within the R-matrix revealed the direction and extent of confusions among syntactic structures.

Results and Discussion

Degree of stimulus control

Figure 1 shows the effect of the relative frequencies of the syntactic structures in the speakers' corpora on the discriminative control of responding by the listeners. The dependent variable is the degree of the mismatch between the S- and R- matrices for each group of five Ss, expressed as the sum of the absolute value of the differences between the matrices.

If the listener's assignment of sentences to speakers was entirely under discriminative control, the distribution of structures in the R-matrix would be the same as that in the S-matrix and the sum of their absolute differences would be zero over all S-matrices. This hypothetical outcome is shown in Fig. 1 by the line labelled "probability matching". A second possible finding, labelled "random assignment", would be obtained if the listeners' assignments were not under discriminative control at all. The third hypothetical outcome, labelled "maximizing", would be obtained if the listeners assigned all 36 sentences with a given structure to the speaker who said that structure most frequently.

The obtained error function, which compares the first S-matrix presented to each group with the mean of the corresponding five R-matrices, is non-monotonic and peaks at stimulus matrices of intermediate uncertainty. Comparison of the obtained and hypothetical functions suggests that listeners assigned all occurrences of a given structure to the speaker for whom that structure was dominant, when the dominance in frequency for one structure was marked. However, when the relative frequencies of the four syntactic structures in the S-matrix were more nearly equal, discriminative control was minimal and the structures in the R-matrix were distributed essentially at random. Examination of the other nine R-matrices for each listener supports this generalization, although order effects somewhat confound the analysis.

It is interesting to note that the transition from maximizing to random assignment occurred near the point of intersection of the two hypothetical error functions. Had the dominant structure continued to pre-empt discriminative control, a greater mismatch between the S- and R- matrices than that produced by random assignment would have resulted.

Overlap in stimulus control

Two or more syntactic forms in the corpus of one speaker were classified as "confused" if the listener assigned ten or more sentences of each form to that speaker. For example, if speaker 15 read 27 P-sentences and three each of D-, N-, and Q- sentences and a listener made the assignments: 20P, 15D, 4N, 4Q, then P and D are said to be confused when P was the dominant structure.

Figure 2 shows the pattern of confusions when D, N, P, and Q were the dominant structures. When D was dominant, 50 per cent of the confusions that occurred were associated with D and P. Similarly, when P was dominant, most of the confusions occurred with D and P. When N was dominant, the greatest number of confusions still occurred between D and P, although when Q was dominant, Q generalized with P most often.

Table 2, which summarizes these findings, shows the total number of confusions associated with each syntactic structure as a function of the dominant structure in the speaker's corpus. As might be expected, most of the confusions that occurred in assigning sentences to a speaker involved the syntactic structure read most often by the speaker. It is interesting to note, however, that P-sentences were confused with all other structures as often as D-sentences when D was dominant and, likewise, there are almost as many D confusions as there are P-confusions when P was dominant.

The column marginals of the percent-confusion table show that nearly twice as many confusions occurred when D or P was dominant in a speaker's corpus than when N or Q was dominant. Collapsing over the dominant structures read by the speakers, the row marginals show that D or P sentences were more often confused with the other three structures, that is, more often assigned along with them to the same speaker, than were the N- or Q- sentences. Q-sentences were least often confused with the other structures.

Summary

Forty scripts were prepared that contained the same 36 kernel sentences with one of four syntactic structures: declarative, negative, passive, or querry. The relative frequency of the structures in each script was varied systematically. Tape recordings of the scripts, read by forty speakers, were presented to listeners who sorted a (different) set of 144 sentences according to their estimate of which speaker was most likely to have said each sentence. The distribution of syntactic structures in the stimulus scripts was compared with the corresponding distribution arrived at by the listener.

When a particular syntactic structure predominated in a speaker's corpus, the listeners tended to assign all the sentences of the dominant form exclusively to that speaker. When the relative frequencies of the syntactic structures in a corpus were more nearly equal, the listeners tended to assign structures at random. Most of the confusions in discrimination, that is, assignment of a structure to a speaker much more often than it was uttered, were associated with the declarative and passive sentence structures. These syntactic forms generalized most often with each other and more often with the remaining syntactic forms than either negative or interrogative sentences.

Reference

Miller, G. A. Some psychological studies of grammar. Amer. Psychol., 1962, 17, 748-762.

Table 1

FREQUENCIES OF THE SYNTACTIC STRUCTURES READ BY EACH SPEAKER

Matrix No.	H (Bits)	Speaker No.	Dominant Structure	Frequency			
				D	N	P	Q
1	2.0	1	D	36	0	0	0
		2	N	0	36	0	0
		3	P	0	0	36	0
		4	Q	0	0	0	36
2	2.6	5	D	33	1	1	1
		6	N	1	33	1	1
		7	P	1	1	33	1
		8	Q	1	1	1	33
3	2.9	9	D	30	2	2	2
		10	N	2	30	2	2
		11	P	2	2	30	2
		12	Q	2	2	2	30
4	3.2	13	D	27	3	3	3
		14	N	3	27	3	3
		15	P	3	3	27	3
		16	Q	3	3	3	27
5	3.5	17	D	24	4	4	4
		18	N	4	24	4	4
		19	P	4	4	24	4
		20	Q	4	4	4	24
6	3.6	21	D	21	5	5	5
		22	N	5	21	5	5
		23	P	5	5	21	5
		24	Q	5	5	5	21
7	3.8	25	D	18	6	6	6
		26	N	6	18	6	6
		27	P	6	6	18	6
		28	Q	6	6	6	18
8	3.9	29	D	15	7	7	7
		30	N	7	15	7	7
		31	P	7	7	15	7
		32	Q	7	7	7	15
9	3.98	33	D	12	8	8	8
		34	N	8	12	8	8
		35	P	8	8	12	8
		36	Q	8	8	8	12
10	4.0	37	D	9	9	9	9
		38	N	9	9	9	9
		39	P	9	9	9	9
		40	Q	9	9	9	9

Table 2

The syntactic structures which were confused when a dominant structure was sent.

FREQUENCY OF CONFUSION

		Dominant Structure			
		D	N	P	Q
	D	74	28	74	23
Structures	N	35	27	46	22
Confused	P	68	27	84	28
	Q	23	20	36	31

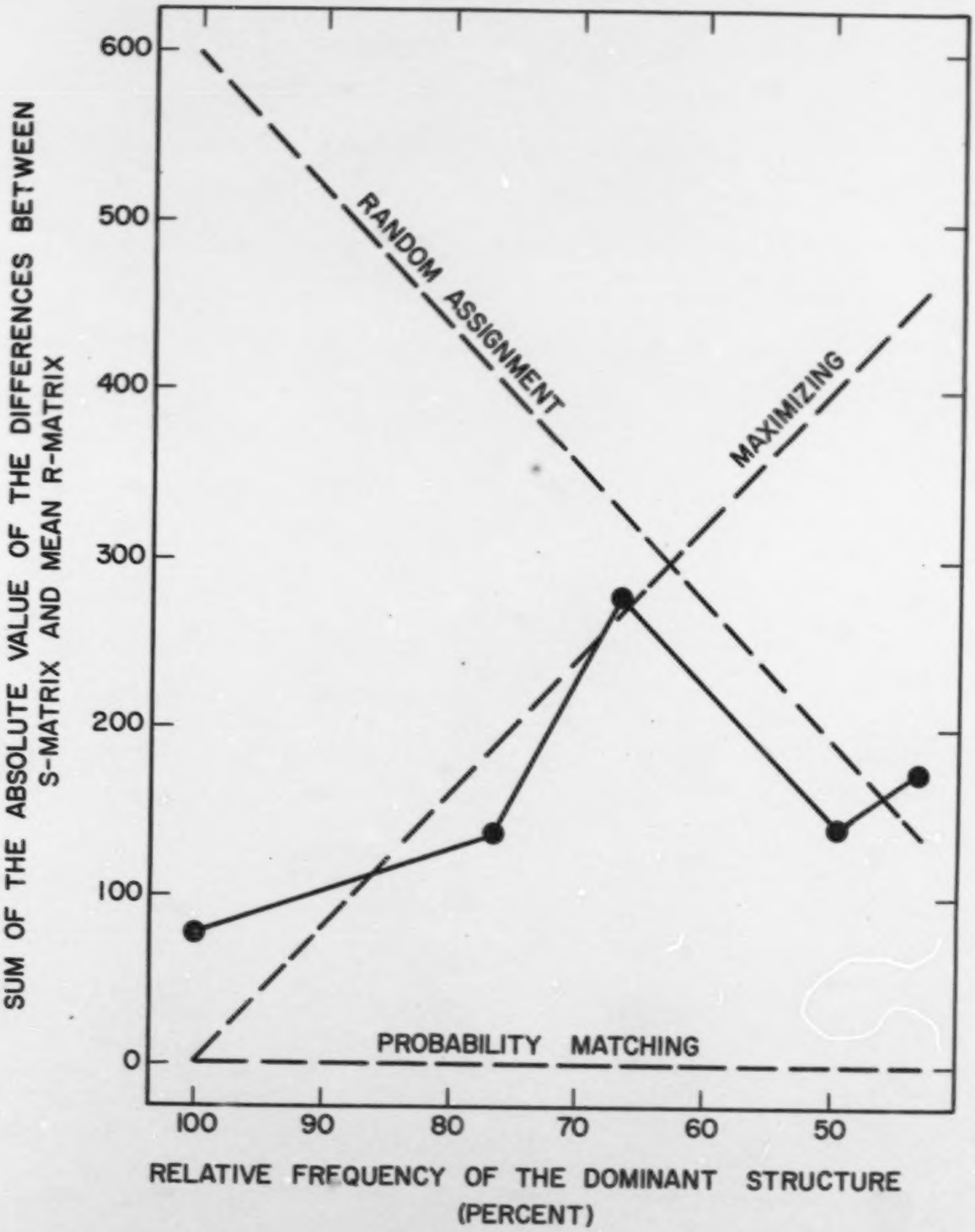
PER CENT CONFUSION

		Dominant Structure				
		D	N	P	Q	TOTAL
	D	12	4	12	4	31
Structures	N	5	4	7	3	20
Confused	P	11	4	13	4	32
	Q	4	3	6	5	17
	TOTAL	31	16	37	16	

Figure Captions

Fig. 1. The effect of the relative frequencies of four syntactic structures in a speaker's corpus on the accuracy of discriminative responding by a listener. After hearing a set of four speakers each read 36 sentences, with the syntactic structures: declarative, negative, passive, and querry occurring in different frequencies, each of five listeners sorted a comparable set of 144 sentences according to his estimate of which speaker was most likely to have said each sentence. The difference between the distributions of structures in the stimulus matrix and in the corresponding average response matrix is plotted as a function of the relative frequency of the dominant syntactic structure in the stimulus matrix. The dotted lines indicate hypothetical results (see text).

Fig. 2. Patterns of confusion in the discrimination of syntactic structure. The frequency of a confusion between two or more structures is the number of occassions on which 10 or more sentences of each structure were assigned to the same speaker. The dominant structure is the syntactic form uttered most frequently by the speaker when the confusion occurred.



SUM OF THE ABSOLUTE VALUE OF THE DIFFERENCES BETWEEN S-MATRIX AND MEAN R-MATRIX
600
500
400
300
200
100
0
RANDOM ASSIGNMENT
MAXIMIZING
PROBABILITY MATCHING
100
90
80
70
60
50
RELATIVE FREQUENCY OF THE DOMINANT STRUCTURE (PERCENT)

Fig. 1

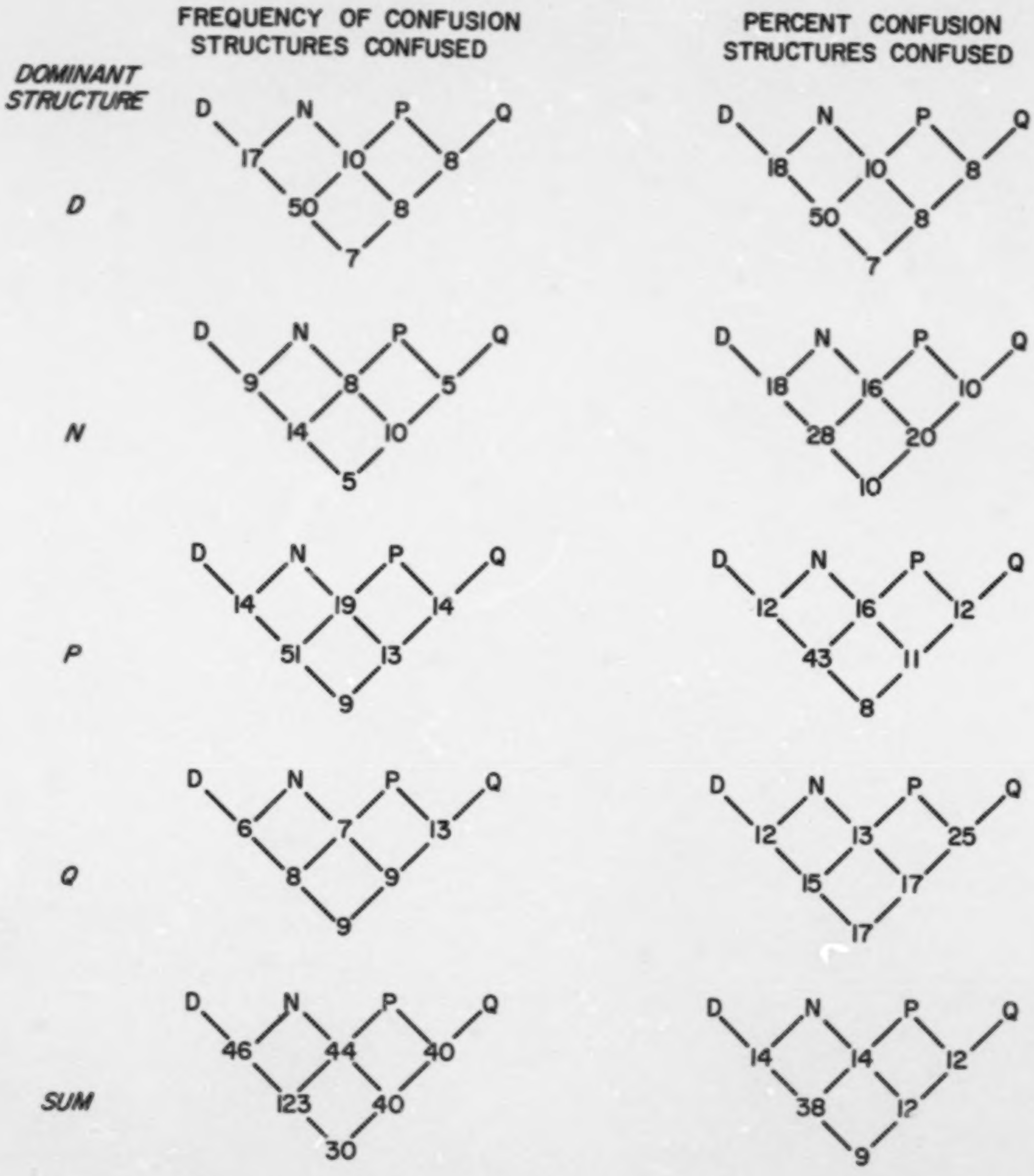

FREQUENCY OF CONFUSION
STRUCTURES CONFUSED
PERCENT CONFUSION
STRUCTURES CONFUSED
DOMINANT STRUCTURE
D
N
P
Q
SUM
D N P Q 17 10 8 50 8 7
D N P Q 18 10 8 50 8 7
D N P Q 9 8 5 14 10 5
D N P Q 18 16 10 28 20 10
D N P Q 14 19 14 51 13 9
D N P Q 12 16 12 43 11 8
D N P Q 6 7 13 8 9 9
D N P Q 12 13 25 15 17 17
D N P Q 46 44 40 123 40 30
D N P Q 14 14 12 38 12 9

Fig. 2

RATIO SCALES, CATEGORY SCALES, AND VARIABILITY IN THE PRODUCTION OF LOUDNESS AND SOFTNESS

Bruce Schneider
The University of Michigan

Ratio Scales, Category Scales, and Variability in the Production of Loudness and Softness

Bruce Schneider
Behavior Analysis Laboratory
University of Michigan[1]

A considerable number of psychophysical studies have shown that category scales are non-linearly related to ratio scales of subjective magnitude on prothetic continua (Stevens, J. C. & Mack, J. D., 1959; Stevens, S. S. & Galanter, E. H., 1957; Stevens, S. S. & Geraldine Stone, 1959; Torgerson, W. S., 1960). Stevens and Guirao (1962) provide the most recent account of this observed departure from linearity in terms of a variability model. This model involves the assumption that ratio-scaling procedures (most often magnitude estimation and magnitude production) provide a valid measure of the psychological continuum; the model then seeks to explain the non-linear relation between category and ratio scales in terms of the variability associated with judgments along the continuum. According to Stevens and Guirao, variability is not constant along the psychological continuum, but increases with increasing magnitude. Consequently, when an observer attempts to assign stimuli to categories equally spaced in psychological units, he makes few errors at lesser psychological magnitudes, where he can easily discriminate two points that are separated by a given distance, and he makes many errors at higher magnitudes, where two points, although separated by the same distance, are not so easily discriminated, and hence are assigned more often to the same category. The result of this bias is that the category scale is approximately linearly related to the magnitude scale for lower values of psychological magnitude; however, at higher magnitudes, confusion increases and the points are grouped more often in the same category. The graphic outcome is that the category scale appears concave downward when plotted against the magnitude scale (Stevens & Galanter, 1957).

This article examines the variability model in the light of a precise description of the empirical relations that enter into it: the ratio scale of subjective magnitude, the corresponding category scale, and the variability of judgments in both physical and psychological units. These relations are determined quantitatively, through repeated measurement with a single observer, for the psychological continuum, loudness, and its inverse, softness. The fit of the variability model to findings provided by the new technique of inverse scaling is assessed.

Method

Procedure

Our subject, who began as an untrained observer, gave judgments of sensory magnitude according to four procedures: 1) magnitude production of loudness; 2) category production of loudness; 3) magnitude production of softness; 4) category production of softness. The corresponding sets of instructions, read to S before each session, were as follows. (The words in the parentheses were used when appropriate.)

Magnitude production. "This is an experiment to see how you perceive the loudness (softness) of sounds. When the experiment begins, I will present a standard sound to you. Call the loudness (softness) of this sound '10'. I will then present a series of numbers, one at a time. Your task will be to adjust the noise so that its loudness (softness) stands in the same proportion to the loudness (softness) of the standard noise as the number

I have given you stands to 10. For example, if I give you the number 20, you should adjust the noise so that it is twice as loud (soft) as the standard. If I give you the number 3.3, you should adjust the noise so that it is one-third as loud (soft) as the standard. Always adjust the noise so that its loudness (softness) stands in the same proportion to the standard as the number I have given you stands to 10."

Category production. "In this part of the experiment we are again interested in seeing how you perceive the loudness (softness) of sounds. Before the experiment begins, I will present two noises in succession over the earphones. The loudness (softness) of the first will be called 'one', and the loudness (softness) of the second will be called 'five'. I will then present a series of numbers from one to five in irregular order. You should regard these numbers as marking off equal distances in loudness (softness). Your task is to adjust the noise so that its loudness (softness) corresponds to the number I have given you."

In the magnitude production sessions, the standard loudness had a sound pressure of 1.78 microbars (79 db, sound pressure level); it was presented and identified at the start of the stimulus series and again after every 25th determination. After a criterion value (2.5, 5, 10, 20, 40) was presented and S was satisfied with his production, he pushed a button which terminated the noise. One hundred and forty determinations were obtained for each magnitude in irregular order; they were collected in five sessions, lasting approximately two hours each.

In the category production sessions, the sound pressures defining the end categories were the geometric means of the sound pressures assigned to the lowest and highest values in the corresponding magnitude production experiments (.31 and 17.6 microbars for loudness; 10.5 and .275 microbars for softness). These levels were also presented and identified at the outset and after every 25th determination. Four sessions were employed.

Apparatus

In all four experimental conditions, S was seated in a sound-treated room; he wore a binaural headset with calibrated TDH 39 earphones, matched in frequency response, and mounted in sponge neoprene cushions. The stimuli were produced by a noise generator (Grason-Stadler 901A) whose spectrum was constant in amplitude up to 1,000 cps and fell at 12 db/octave thereafter. The signal was sent to a linear potentiometer, adjusted by S. In order to minimize bias in these productions arising from the position of the potentiometer on prior trials, an attenuator (Hewlett-Packard 350D) was interposed between the potentiometer output and the earphones, and E varied the starting level of the noise from trial to trial. Furthermore, S was required to set the potentiometer to the zero position after each trial. The subject could adjust the level of the noise for the appropriate magnitude or category production over the following ranges: 0- 75.4, 0- 30.4, 0- 8.1, 0- 3.4, and 0- 1.7 microbars. The voltage produced at the earphones was measured by a vacuum tube voltmeter (Hewlett-Packard 400D) and later converted to sound pressure by means of earphone calibrations with equivalent sine-wave voltages.

Results and Discussion

Magnitude scales

Figure 1 shows the geometric means of the sound pressures produced to each of the criterion values for magnitude production of loudness. The determinations in sessions 1 through 5 are well fit in each case by a straight line in log-log coordinates, in other words, a power function. The slope (exponent) of this function, determined by the method of least squares, ranges from .65 to .75. The value of the slope for sessions 1-5 combined, 0.70, agrees with that obtained by Stevens and Guirao (1962) for 2 determinations on each of 10 *S*s. There is no evidence in the present study of a systematic change in the slope or intercept of the loudness function over the course of the five sessions.

Figure 2 shows that magnitude production of softness is also well described by a power law. The range of exponents (slopes) of this function is, however, greater than that for loudness: -.71 to -.90. The softness function becomes somewhat steeper in the first 4 sessions, but this trend is reversed in session 5. Similarly, the intercept increases in sessions 1-4, reverts in session 5. For the combined sessions, the softness function (Fig. 2f), with exponent -.76, is very nearly the reciprocal of the loudness function (Fig. 1f), with exponent +.70.

Category scales

Category production scales have been shown to fall somewhere between a logarithmic function and a power function of stimulus intensity (Stevens, 1960). This finding is confirmed by the results of category productions of loudness and softness, shown in Figs. 3 and 4, respectively. For loudness, the curves are slightly concave upward, for softness, the curves are slightly concave downward; in either case the category scale does not describe a straight line in the semi-log coordinates.

In order to arrive at an equation describing the category scales, we observe that, for both loudness and softness, the departure from a logarithmic function occurs for low values of sound pressure (ϕ). This suggests that a constant added to the mean sound pressure of a category would produce a straight line function when the means are plotted against log $(\phi + C)$. Accordingly, equations of the form

$$G_L = b_1 \log_{10}(\phi + c) + a_1$$

$$G_S = b_2 \log_{10}(\phi + c) + a_2$$

where,
G_L = category loudness
G_S = category softness
ϕ = sound pressure (microbars)
c = a transformation constant
a_1, a_2, b_1, b_2, empirical constants to be determined from the data

were fit to the data (method of least squares). The same constant, c = .0427 microbars, was used for both loudness and softness. The slope constants for these functions turned out to be 3.1 for loudness and -3.8 for softness. In other words, the category scale for loudness is approximately the complement of the category scale for softness. The fit of the equations to the obtained points (plotted in log ϕ coordinates) is shown for sessions 1-4 combined in Fig. 3e (loudness) and Fig. 4e (softness). These equations also describe the category scales of loudness and softness of a 1,000 cps tone, obtained by Stevens and Guirao (1962) with a group of observers (Fig. 5). The curves fit to their data were obtained in the manner described above, after adding the constant c = .09 to each of the mean sound pressures. The two functions are nearly complementary; the slope for loudness equals 4.2 and the slope for softness, -3.8.

The findings discussed so far provide the following description of ratio scales and category scales for loudness and softness. Magnitude productions of loudness and softness yield reciprocal scales that grow as a power function of sound pressure (ϕ). Category productions of loudness and softness yield complementary scales that grow as a logarithmic function of sound pressure plus a constant $(\phi + c)$. In mathematical form and with the parameters determined in this experiment:

$M_L = 5.5\ \emptyset^{0.7}$ M_L = loudness magnitude

$M_S = 15.8\ \emptyset^{-0.76}$ M_S = softness magnitude

$G_L = 3.1 \log(\emptyset + .4427) + 1.3$

$G_S = -3.8 \log(\emptyset + .4427) + 4.9$

Category scales vs magnitude scales

When we plot the category scale as a function of the magnitude scale for loudness (Fig. 3f) and for softness (Fig. 4f) we obtain a curious result. In linear coordinates both scales are concave downward; however, in semi-log coordinates, the scale for loudness is concave upward, that for softness concave downward. It follows that the degree of concavity, in linear coordinates, is greater for the softness function than it is for the loudness function. The logarithmic transformation on the abscissa over-corrects the downward concavity of the loudness scale and leaves the function slightly concave upward. This same transformation on the softness scale does not straighten the function, which still remains slightly concave downward.

The differing concavities of the softness and loudness category scales, relative to their corresponding ratio scales, turns out to be a necessary consequence of the empirical form of these four scales, which were described in a previous section of the paper. Using the equations stated earlier, we may derive an expression for the category scale as a function of psychological magnitude. Specifically:

$M_L = k_1\emptyset^n, \quad \log M_L = n \log \emptyset + \log k_1,$

$\log \emptyset = (1/n) \log (M_L/k_1), \text{ and } \emptyset = (M_L/k_1)^{1/n}.$

Similarly, $\emptyset = (M_S/k_2)^{-1/n}$. It was shown earlier that

$G_L = b \log (\emptyset + c) + a_1$ and $G_S = -b \log (\emptyset + c) + a_2$.

Let $k_1' = (1/k_1)^{1/n}$, $k_2' = (1/k_2)^{-1/n}$. Substituting for ϕ in G_L, G_S:

$G_L = b \log (k_1' M^{1/n} + c) + a_1$ and $G_S = -b \log (k_2' M^{-1/n} + c) + a_2$.

In terms of the parameters of the category and ratio scales shown earlier:

$$G_L = 3.1 \log (.11M^{1/.7} + .4427) + 1.3$$

$$G_S = -3.8 \log (38.29M^{-1/.76} + .4427) + 4.88.$$

These functions, relating category to ratio scales of loudness and softness, are shown in Figs. 3f and 4f along with the obtained data points. It can be shown (Appendix A) that G_L is always concave upward and G_S concave downward (for $c > 0$) in logarithmic coordinates. In linear coordinates, G_S must be more concave downward than G_L, that is, the rate at which its sense of concavity changes must be greater (Appendix B).

(In the special case when $c = 0$, the category scales are logarithmic functions of the corresponding physical continua. Although this is not a likely empirical finding, we should note that, in this case,

$$G_L = b \log (M^{1/n}) + a_1 = b/n \log M + a_1$$

$$G_S = -b \log (M^{-1/n}) + a_2 = b/n \log M + a_2$$

and the functions do not differ in concavity.)

When the expressions relating category and magnitude scales of softness and loudness are applied to the group data collected by Stevens and Guirao (Fig. 6), we find an equally good fit, and confirmation of the conclusion that the concavities of these two functions necessarily differ.

Eisler (1962) has reported findings that appear, at first, to contradict this conclusion. In three series of experiments in which *Ss* gave magnitude and category estimates of the loudness and softness of noise, the functions, relating category scales to log magnitude were found to be concave upward for *both* loudness and softness. However, an examination of the functions relating psychological magnitude (magnitude estimation) to sound pressure (Fig. 7) reveals a rather curious anomaly. Although a power function describes the growth of estimates of loudness, it is evident that a reciprocal function does not describe the growth of softness. In fact, the estimates of softness collected by Eisler do not follow a power law at all. It is not surprising, therefore, that the predicted difference in concavity did not obtain in his experiment. In fact, as long as the ratio scales of loudness and softness are reciprocal power functions, and the category scales are complementary logarithmic functions (of $\phi + c$), the functions relating category to magnitude scales will differ in concavity.

The variability model

The variability model relates the degree of concavity of the category scale (when plotted against psychological magnitude) to the increase in variability with increasing psychological magnitude. The arguments presented above show that it is necessarily true that this concavity is greater for softness than it is for loudness. If the variability hypothesis is to be in accord with the necessary outcome, then one or both of two conditions must obtain: 1) psychological variability is greater for softness than it is for loudness, 2) psychological variability grows at a faster rate for softness than for loudness. In this section we examine variability in the magnitude and category production of loudness and of softness.

Variability in physical units. Figure 8 presents the distribution of magnitude or category productions (in physical units) observed at each criterion value in both loudness and softness scaling. The 550 determinations under each procedure are shown; the data for the initial sessions were omitted in order to exclude atypical measures of variability. The distributions are approximately normal in all four experimental conditions. Variability increases with increasing sound pressure for both loudness and softness scaling. Since loudness also increases as a function of stimulus magnitude, this means that the variability of loudness productions, measured in physical units, increases as a function of loudness. However, softness decreases as a function of sound pressure, while the variability of softness productions, measured in physical units, increases as a function of sound pressure. Hence, the variability of softness decreases as a function of softness. We cannot, then, attribute the difference in the degree of concavity of the loudness and softness scales (Figs. 3f, 4f) to the way in which variability, measured in physical units, grows as a function of psychological magnitude; in one case it increases, in the other it decreases.

Furthermore, the rate at which variability changes as a function of physical magnitude is the same for both the loudness and softness scales. Figure 9 shows the standard deviations of the distributions for loudness (filled circles) and softness (unfilled circles) plotted against the physical magnitude of their means. It will be seen that the standard deviation is approximately the same linear function of sound pressure for both the loudness and softness scales, obtained by category or by magnitude production. Since variability grows, in physical units, at the same rate for both loudness and softness, we may conclude that, on this second count, as well, the findings for variability in physical units do not explain the different concavities of the loudness and softness category scales relative to their magnitude scales.

Psychological variability. Figure 10 presents the "psychological variability" associated with the production of a particular magnitude or category. To obtain these distributions, each production was transformed into psychological units based on the reciprocal scales obtained from the magnitude production experiments. A best estimate of the absolute value of the exponent for the softness and loudness scales may be /.73/, which is intermediate between the observed exponents of /.70/ (loudness) and /.76/ (softness). Accordingly, each loudness production was transformed to psychological units by $M_L = 5.5\phi^{0.73}$ and each softness production by $M_S = 15.8\phi^{-0.73}$. (The initial sessions again were omitted.)

In psychological units, variability increases with increasing magnitude for both loudness and softness. The frequency distributions, although slightly skewed, again approximate normal curves. If the variability hypothesis is to account for the differing concavities of the loudness and softness scales, it is necessary that variability be either greater or grow more rapidly for softness than for loudness. Figure 11 presents the standard deviation of magnitude productions in psychological units as a function of the mean psychological magnitude for both scales. When a straight line was fit to the loudness and softness points separately (by the method of least squares) a small difference in the slopes of the lines, 0.03, suggested that indeed the variability of softness productions grew more rapidly, however slightly, than the variability of loudness productions. This inference was not borne out by an F-test of the ratio of the variances at several corresponding points along the least-squares lines. In fact, the line of best fit to the softness data at no point fell within the region ($\alpha = .01$) which permitted rejection of the null hypothesis that the softness and loudness variances were sampled from the same population. Since the variability of loudness and of softness productions grows at approximately the same rate as a function of psychological magnitude, the variability model fails to account for the difference in concavity between the loudness and softness category scales relative to their magnitude scales.

Summary

Several studies have shown that category scales are non-linearly related to ratio scales of subjective magnitude. To account for this departure from linearity, a variability model has been proposed by Stevens and Guirao. This article examines the model in the light of a precise description of the empirical relations that enter into it: the ratio scale of subjective magnitude, the corresponding category scale, and the variability of judgments in both physical and psychological units. These relations are determined quantitatively, through repeated measurement with a single observer, for the psychological continuum, loudness, and its inverse, softness. The ratio scales were shown to be reciprocals, and the category scales complements. The category scale of softness was found to be more concave downward relative to its magnitude scale than was the category scale of loudness. This outcome was shown to be a necessary consequence of the empirical equations relating the four scales to physical magnitude. Variability in physical units was found to increase with increasing stimulus magnitude at the same rate for both loudness and softness productions. In psychological units, variability increased with increasing psychological magnitude. The psychological variability associated with loudness productions did not differ systematically from that for softness productions. Hence, the variability model was found not to accord with the observed difference in concavity between softness and loudness category scales relative to their respective psychological magnitude scales.

References

Eisler, H. Empirical test of a model relating magnitude and category scales. *Scand. J. Psychol.*, 1962, 3, 88-96.

Stevens, J. C. & Mack, J. D. Scales of apparent force. *J. exp. Psychol.*, 1959, 58, 405-413.

Stevens, S. S. On the new psychophysics. *Scand. J. Psychol.*, 1960, 1, 27-35.

Stevens, S. S. & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, 54, 377-411.

Stevens, S. S. & Guirao, M. Loudness, reciprocality, and partition scales. *J. Acoust. Soc. Amer.*, 1962, 34, 1466-1471.

Stevens, S. S. & Stone, Geraldine. Finger span; ratio scale, category scale, and jnd scale. *J. exp. Psychol.*, 1959, 57, 91-95.

Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: theory and practice*. New York: John Wiley & Sons, 1960. Pp. 21-31.

Footnote

1. The guidance of Dr. Harlan Lane, under whose supervision this research was carried out, is gratefully acknowledged.

Appendix A

The direction of concavity of the category scales.

The category scales of loudness and softness are related to their respective magnitude scales by the equations:

$$G_L = f_L(M_L) = b \log (k_1' M_L^{1/n} + c) + a_1$$

$$G_S = f_S(M_S) = -b \log (k_2' M_S^{-1/n} + c) + a_2$$

G_L = category scale, loudness
G_S = category scale, softness
M = psychological magnitude
$a_1, a_2, b, n, k_1', k_2', c$; constants with $b, n, k_1', k_2', c > 0$

In order to test these functions for their sense of concavity in logarithmic coordinates, it is necessary to express G_L, G_S, as functions, g_L, g_S, of log M. (We may drop the subscripts on M_L, M_S since we will treat G_L, G_S, as functions of a single variable, M.) Let $z = \log M$; then, $M = e^z$. Substituting e^z for M in G_L, G_S:

$$G_L = g_L(z) = b \log (k_1' e^{z/n} + c) + a_1$$

$$G_S = g_S(z) = -b \log (k_2' e^{-z/n} + c) + a_2.$$

We can test the sense of concavity of these functions by examining the sign of their second derivatives. If the second derivative is positive over the interval in question, then the curve is concave upward; if it is negative, the curve is concave downward.

The second derivative of G_L is :

$$\frac{d^2G_L}{d\ z^2} = \frac{(b)(1/n)^2\ (k_1')(c)(\log_{10}e)\ e^{z/n}}{(k_1'\ e^{z/n} + c)^2}$$

Since $e^{z/n}$ is always positive and $\log_{10}$ e, n, b, c, and k_1' are positive, the category scale of loudness is necessarily concave upward when plotted against the log of psychological magnitude.

The second derivative of G_S is:

$$\frac{d^2G_S}{d\,z^2} = \frac{-(b)(1/n)^2\,(k_2')(c)(\log_{10} e)\,e^{-z/n}}{(k_2'\,e^{-z/n} + c)^2}\,.$$

Since $e^{-z/n}$ is always positive, and $\log_{10}e$, n, b, c, and k_2' are positive, $\frac{d^2G_S}{d\,z^2}$ is negative; therefore, the category scale of softness is concave downward when plotted against the log of psychological magnitude.

In the special case where $c = 0$, the second derivatives of G_L, G_S are zero, and G_L, G_S are linear functions of the log of psychological magnitude.

Appendix B

The degree of concavity of the category scales.

The category scales of loudness and softness as a function of psychological magnitude are given by the following equations :

$$G_L = b \log (k_1' M^{1/n} + c) + a_1$$

$$G_S = -b \log (k_2' M^{-1/n} + c) + a_2.$$

When G_L and G_S are evaluated with respect to M (linear coordinates) the second derivative of G_S is negative over the entire range of psychological values and that of G_L is also negative except for very small values of M (not considered here).

To determine which of these functions has the greater concavity, we solve for the third derivative of each function, giving the rate of change of concavity. If:

$$\frac{d^3G_S}{dM^3} > \frac{d^3G_L}{dM^3} \qquad (c>0),$$

then the softness function has the greater degree of concavity; i.e., it will appear to curve downward more rapidly than the loudness function.

The third derivatives of G_L and G_S are:

$$\frac{d^3G_L}{dM^3} = \frac{b(\log_{10}e)(1/n)(k_1') M^{1/n} [2(k_1')^2 M^{2/n} - c(1/n + 4)(1/n-1)k_1' M^{1/n} + c^2(1/n-1)(1/n-2)]}{M^3(k_1' M^{1/n} + c)^3}$$

$$\frac{d^3G_S}{dM^3} = \frac{-b(\log_{10}e)(-1/n)(k_2')M^{-1/n} [2(k_2')^2 M^{-2/n} - c(-1/n + 4)(-1/n-1)k_2' M^{-1/n} + c^2(1/n+1)(1/n+2)]}{M^3(k_2' M^{-1/n} + c)^3} .$$

Let us assume that the converse holds, that is

$$\frac{d^3G_L}{dM^3} \geq \frac{d^3G_S}{dM^3} \quad , (c>0).$$

Multiplying both terms of the inequality by $\frac{M^3}{b(\log_{10}e)(1/n)}$, we have

$$\frac{(k_1')M^{1/n}\,[2(k_1')^2\,M^{2/n} - c(1/n+4)(1/n-1)k_1'\,M^{1/n} + c^2(1/n-1)(1/n-2)]}{(k_1'\,M^{1/n} + c)^3}$$

$$\geq \quad \frac{(k_2')M^{-1/n}\,[2(k_2')^2M^{-2/n} - c(-1/n+4)(-1/n-1)k_2'M^{-1/n} + c^2(1/n+1)(1/n+2)]}{(k_2'M^{-1/n} + c)^3}\,.$$

The variable M, psychological magnitude, in the above inequality is related to physical magnitude, ϕ, by :

$$\phi = k_1'M^{1/n} \text{ and } \phi = k_2'M^{-1/n}.$$

Making these substitutions in the above inequality we have;

$$\frac{\phi[2\phi^2-c(1/n+4)(1/n-1)\phi+ c^2(1/n-1)(1/n-2)]}{(\phi + c)^3} \geq \frac{\phi[2\phi^2-c(-1/n+4)(-1/n-1)\phi+c^2(1/n+1)(1/n+2)]}{(\phi + c)^3}\,.$$

Multiplying both terms by $\frac{(\phi + c)^3}{\phi}$, and expanding the parentheses we have,

$$2\phi^2-c\phi/n^2 -3c\phi/n +4c\phi +c^2/n^2 -3c^2/n +2c^2 \geq 2\phi^2 -c\phi/n^2 +3c\phi/n +4c\phi +c^2/n^2 +3c^2/n +2c^2.$$

Cancelling like terms, the inequality reduces to

$$-3c\phi/n -3c^2/n \geq 3c\phi/n + 3c^2/n.$$

Multiplying both terms by $n/3c$ $(c > 0)$ we have

$$-(\phi + c) \geq (\phi + c)\,.$$

But this is a contradiction, since $c > 0$, and ϕ, a physical magnitude, can never be less than zero. It follows that the converse is true; namely, that

$$\frac{d^3G_S}{dM^3} > \frac{d^3G_L}{dM^3} \qquad (c > 0).$$

Therefore, the category scale of softness is necessarily more concave downward than the category scale of loudness. Only in the case where $c = 0$ are the degree of concavities the same. For then,

$$\frac{d^3G_L}{dM^3} = \frac{d^3G_S}{dM^3} = \frac{2(b)(\log_{10}e)(1/n)}{M^3} .$$

Figure Captions

Fig. 1. Magnitude productions of the loudness of white noise for a single observer. Each point in the five sessions (a-e) is the geometric mean of 25-40 determinations. The final figure (f) is a composite of the five sessions; each point is the geometric mean of 140 determinations. The slopes of the straight lines were determined by the method of least squares.

Fig. 2. Magnitude productions of the softness of white noise for a single observer. Each point in the five sessions (a-e) is the geometric mean of 25-40 determinations. The fifth figure (f) is a composite of the four sessions; each point is the geometric mean of 140 determinations.

Fig. 3. Category productions of the loudness of white noise for a single observer. Each point in the four sessions (a-d) is the geometric mean of 25-45 determinations. The fifth figure (e) is a composite of the four sessions; each point is the geometric mean of 140 determinations. The smooth curve is the function of best fit to the obtained data. The final figure (f) presents the category scale of loudness as a function of the magnitude scale (circles) and the logarithm of the magnitude scale (triangles). The smooth curves were determined from the functions relating the category and magnitude scales to sound pressure (see text).

Fig. 4. Category production of the softness of white noise for a single observer. Each point in the four sessions (a-d) is the geometric mean of 25-45 determinations. The fifth figure (e) is a composite of the four sessions; each point is the geometric mean of 140 determinations. The smooth curve is the function of best fit to the obtained data. The final figure (f) presents the category scale of softness as a function of the magnitude scale (circles) and the logarithm of the magnitude scale

(triangles). The smooth curves were determined from the functions relating the category and magnitude scales to sound pressure (see text).

Fig. 5. Category production of loudness (filled circles and softness (unfilled circles) of a tone of 1,000 cps. Each point is the geometric mean of 2 settings by each of 10 observers. The smooth curves are the functions of best fit to the obtained data. After Stevens and Guirao, 1962.

Fig. 6. Category scales of loudness (filled circles) and softness (unfilled circles) as a function of psychological magnitude. The smooth curves were determined from the functions relating the category and magnitude scales to sound pressure (see text). After Stevens & Guirao, 1962.

Fig. 7. Magnitude estimation of loudness (filled circles) and softness (unfilled circles) of white noise. Each point is the geometric mean of 4 estimations by each of 12 observers. After Eisler, 1962.

Fig. 8. Variability, in physical units, for magnitude and category productions of loudness and softness. Each distribution is based on 110 productions. The midpoints of adjacent class intervals are separated by one standard deviation.

Fig. 9. The standard deviation (physical units) of magnitude and category productions of loudness (filled circles) and softness (unfilled circles) as a function of sound pressure. The straight lines were determined by the method of least squares. For clarity, the category production function has been raised three units on the y-axis.

Fig. 10. Variability, in psychological units, for magnitude and category production of loudness and softness. Each distribution is based on 110 productions. The midpoints of adjacent class intervals are separated by one standard deviation.

Fig. 11. The standard deviation (psychological units) of magnitude productions of loudness (filled circles) and softness (unfilled circles) as a function of psychological magnitude. The straight line was determined by the method of least squares.

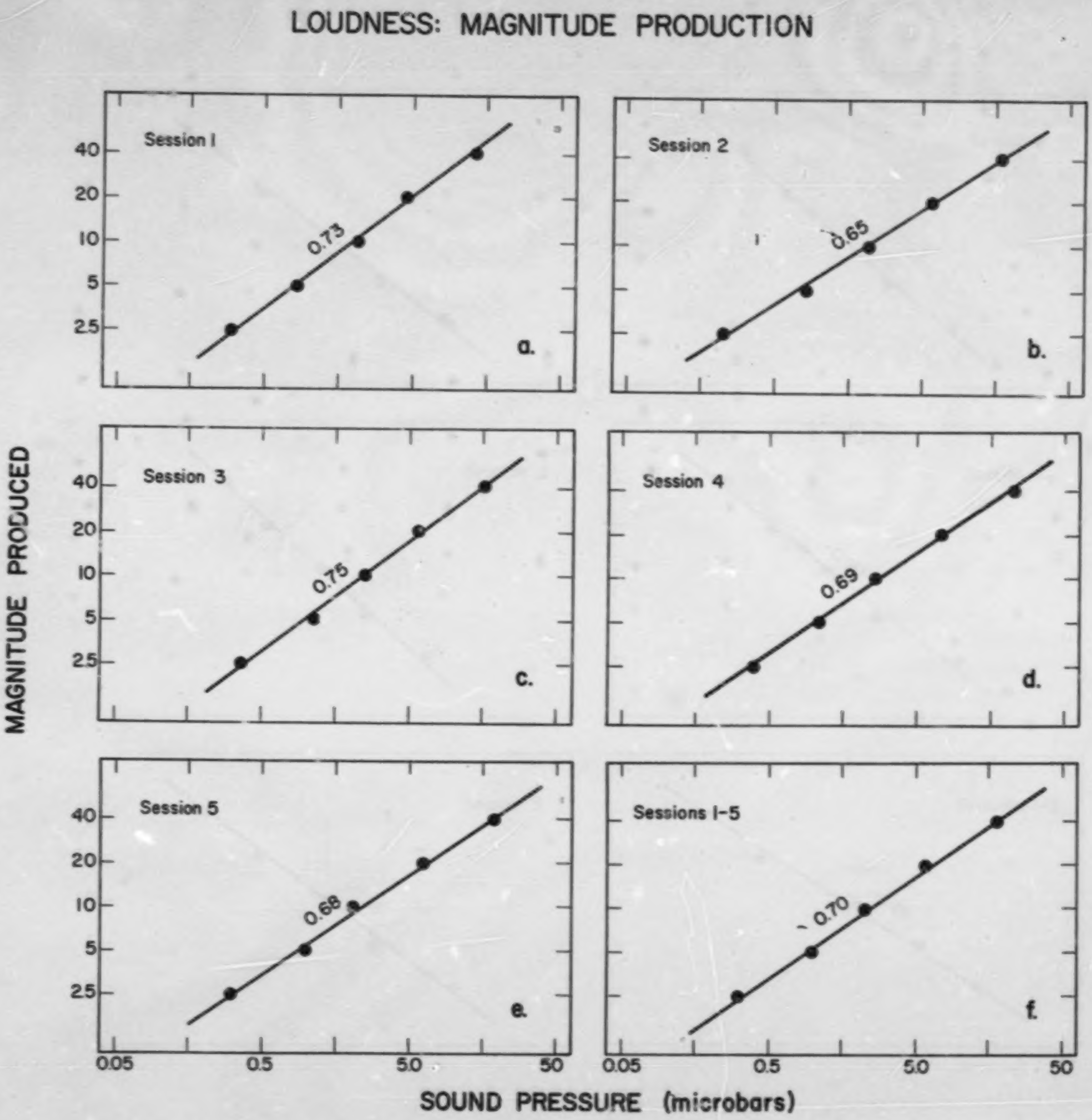

LOUDNESS: MAGNITUDE PRODUCTION
MAGNITUDE PRODUCED
SOUND PRESSURE (microbars)
Session 1
0.73
a.
Session 2
0.65
b.
Session 3
0.75
c.
Session 4
0.69
d.
Session 5
0.68
e.
Sessions 1-5
0.70
f.
40
20
10
5
2.5
0.05
0.5
5.0
50

Fig. 1

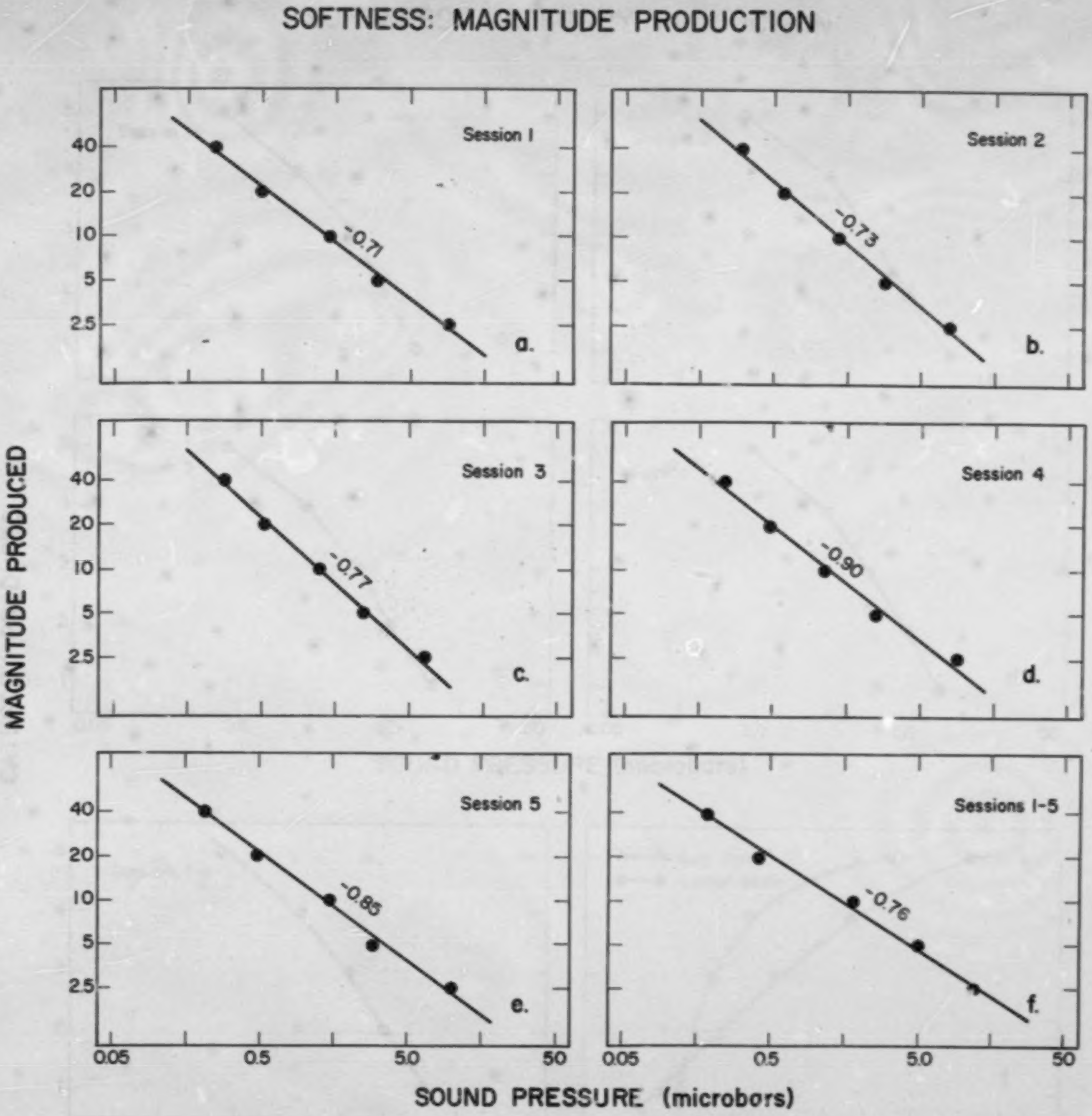

SOFTNESS: MAGNITUDE PRODUCTION
MAGNITUDE PRODUCED
40
20
10
5
2.5
Session 1
-0.71
a.
Session 2
-0.73
b.
Session 3
-0.77
c.
Session 4
-0.90
d.
Session 5
-0.85
e.
Sessions 1-5
-0.76
f.
0.05
0.5
5.0
50
SOUND PRESSURE (microbars)

Fig. 2

SOFTNESS: CATEGORY PRODUCTION

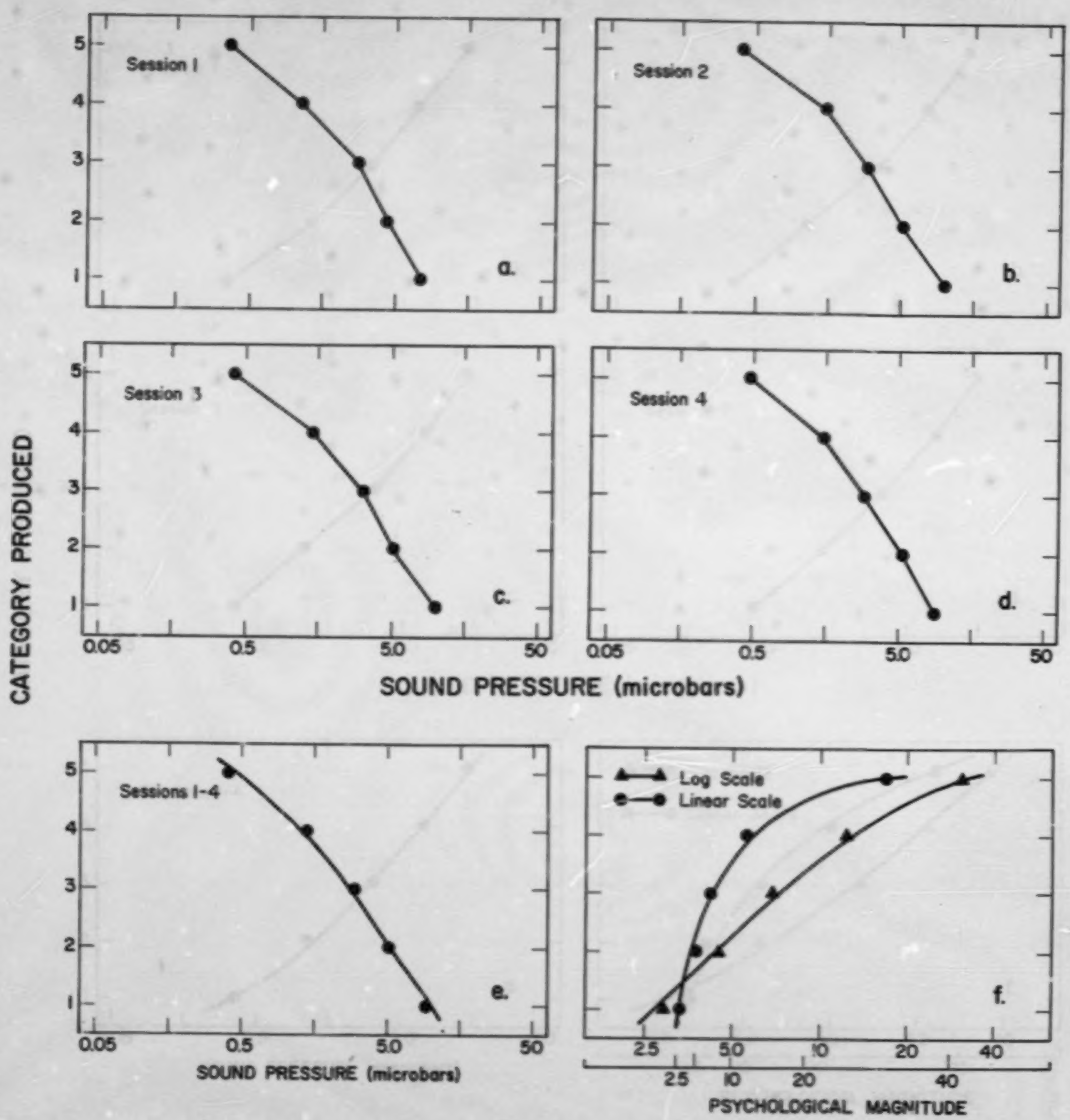


Fig. 3

LOUDNESS: CATEGORY PRODUCTION

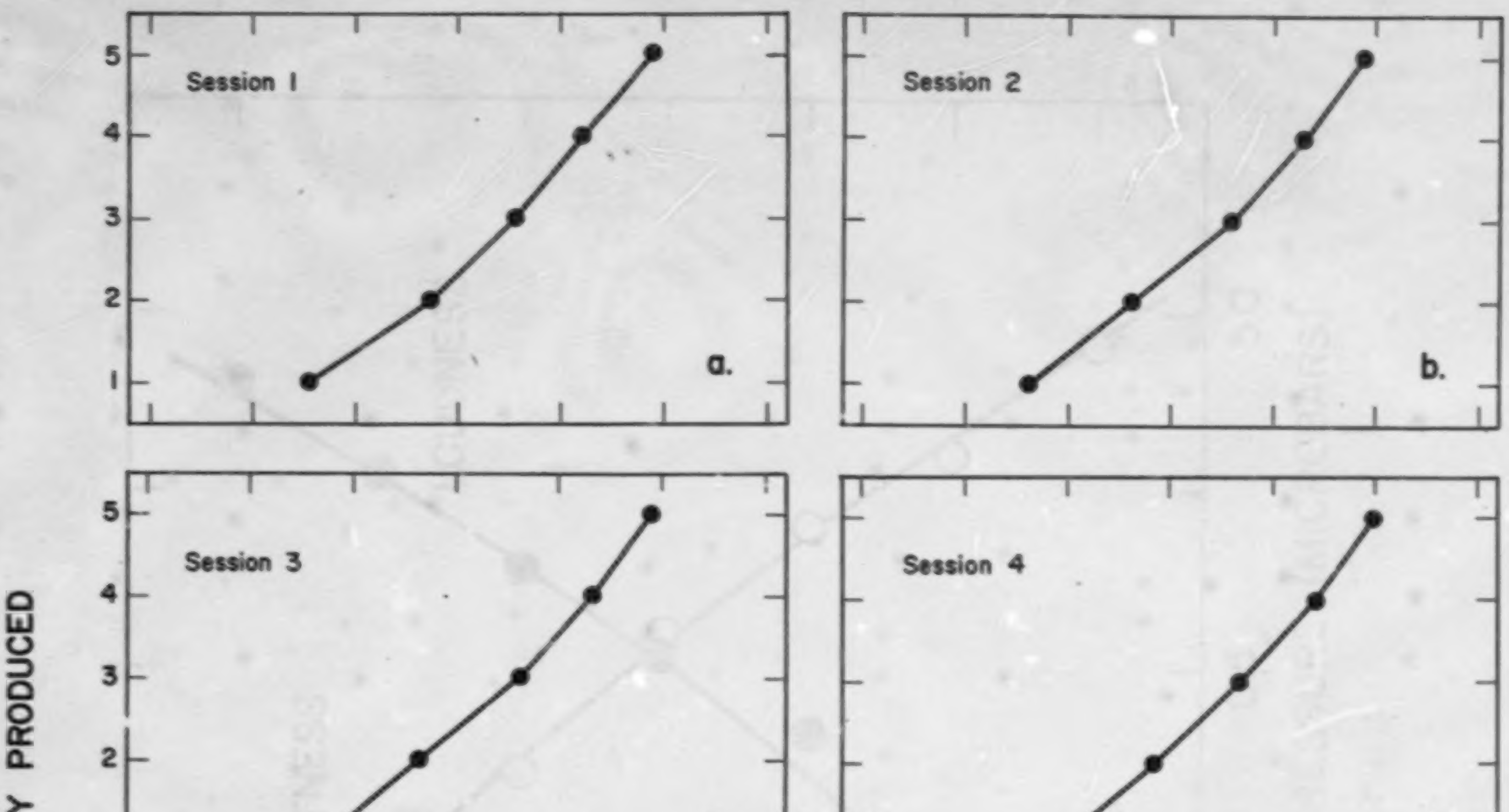


Fig. 4

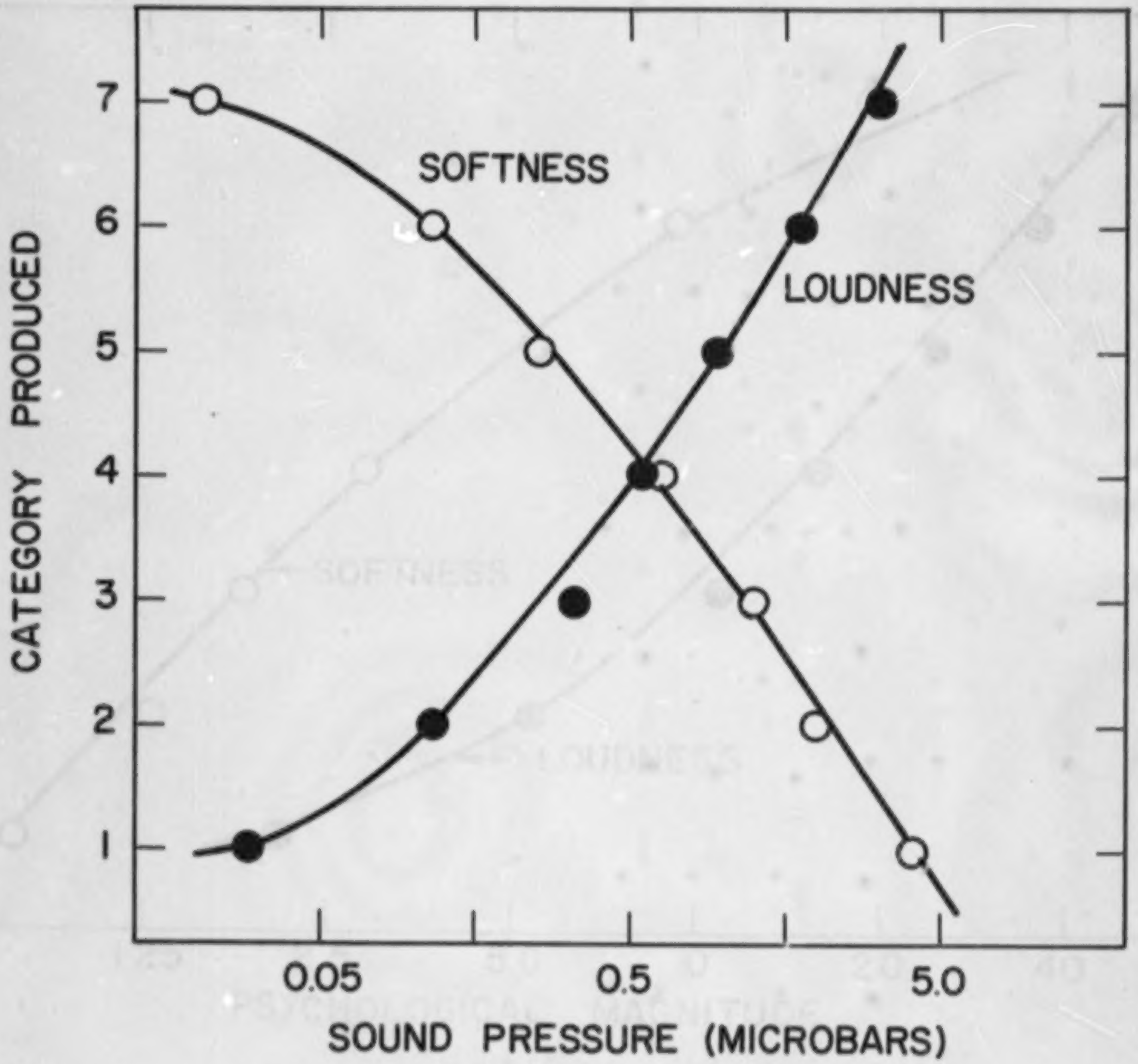

SOFTNESS
LOUDNESS
CATEGORY PRODUCED
7
6
5
4
3
2
1
0.05
0.5
5.0
SOUND PRESSURE (MICROBARS)

Fig. 5

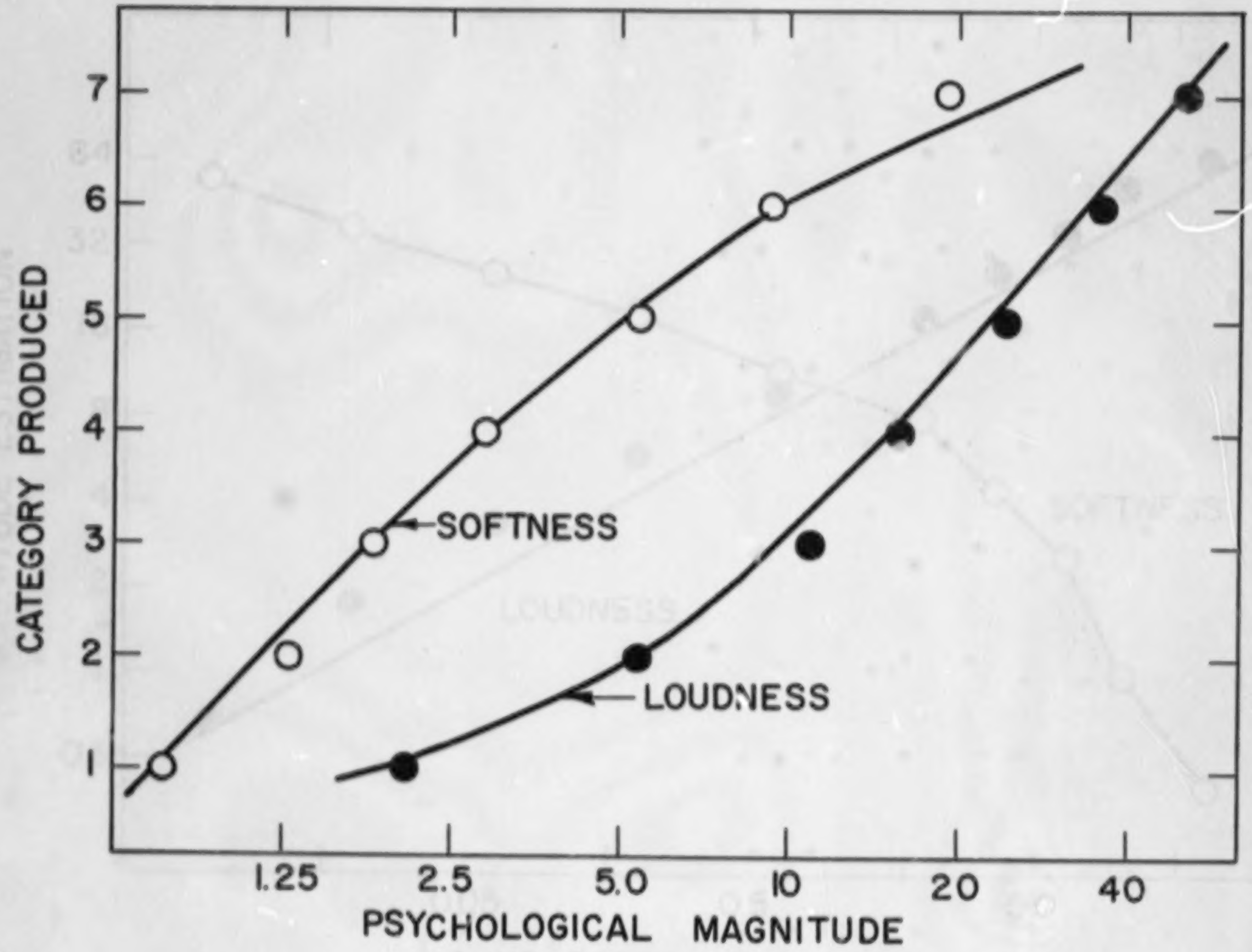

CATEGORY PRODUCED
1
2
3
4
5
6
7
1.25
2.5
5.0
10
20
40
PSYCHOLOGICAL MAGNITUDE
SOFTNESS
LOUDNESS

Fig. 6

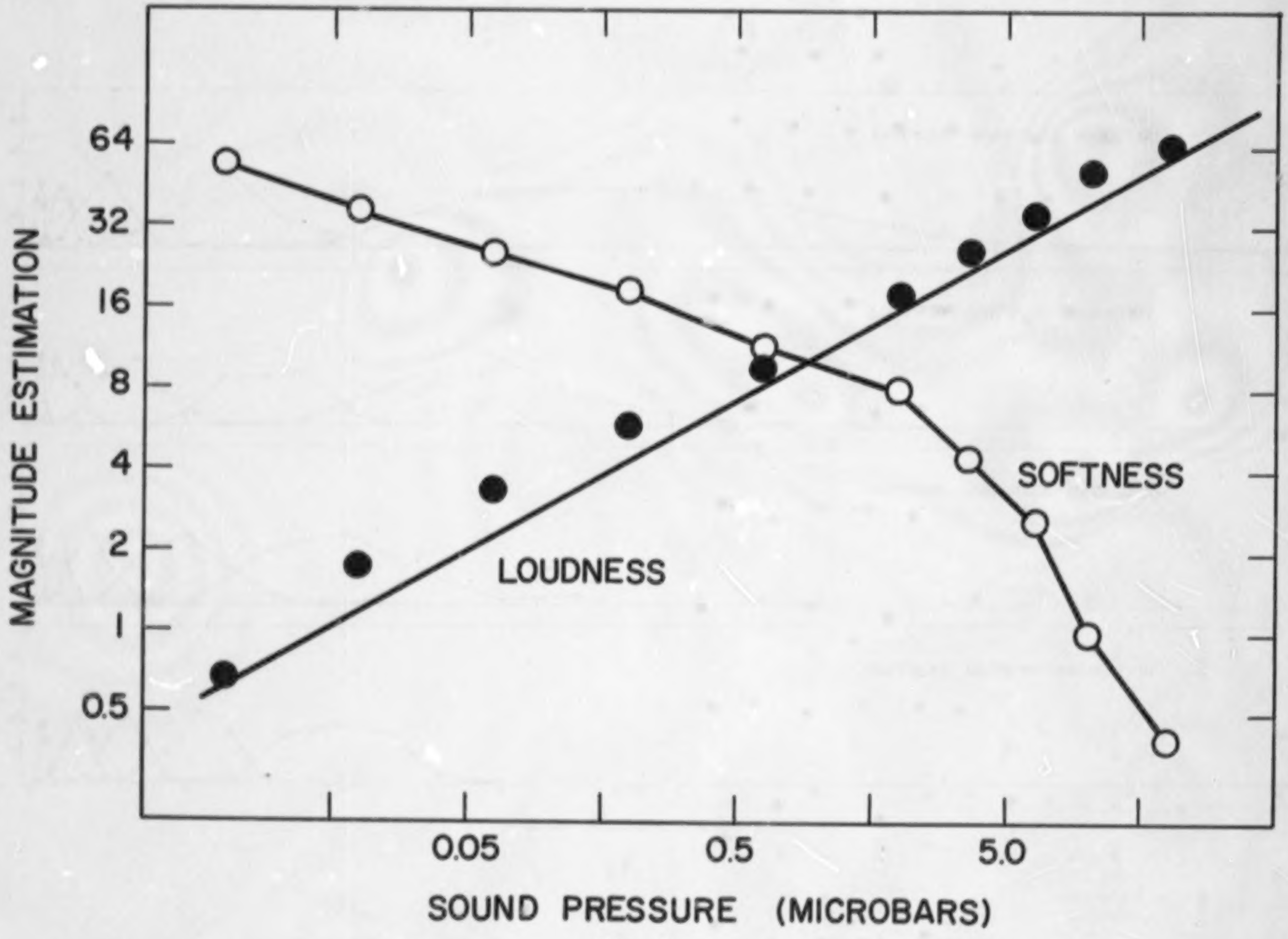

MAGNITUDE ESTIMATION
64
32
16
8
4
2
1
0.5
LOUDNESS
SOFTNESS
0.05
0.5
5.0
SOUND PRESSURE (MICROBARS)

Fig. 7

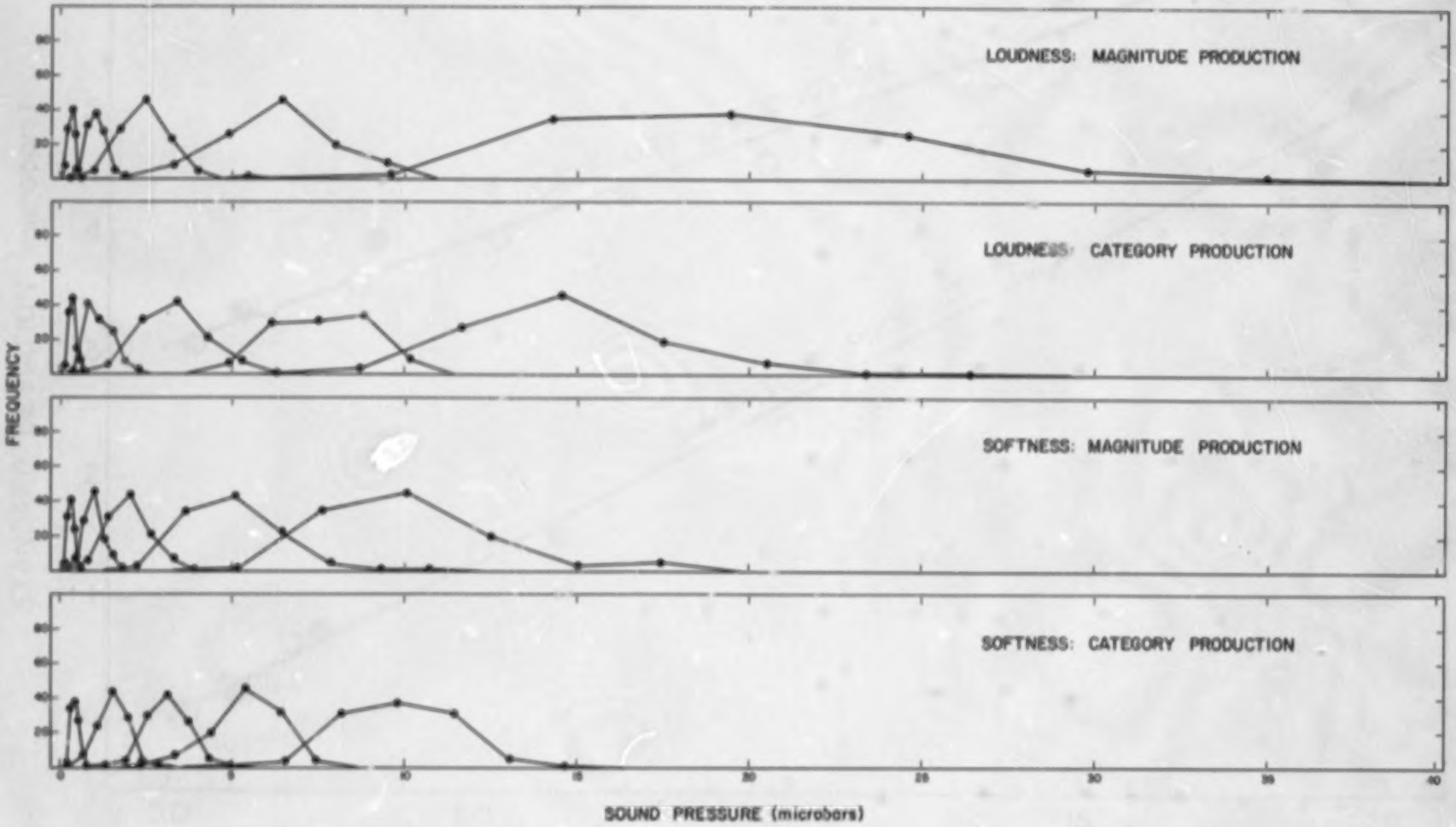

LOUDNESS: MAGNITUDE PRODUCTION
LOUDNESS: CATEGORY PRODUCTION
SOFTNESS: MAGNITUDE PRODUCTION
SOFTNESS: CATEGORY PRODUCTION
FREQUENCY
20
40
60
80
0
5
10
15
20
25
30
35
40
SOUND PRESSURE (microbars)

Fig. 8

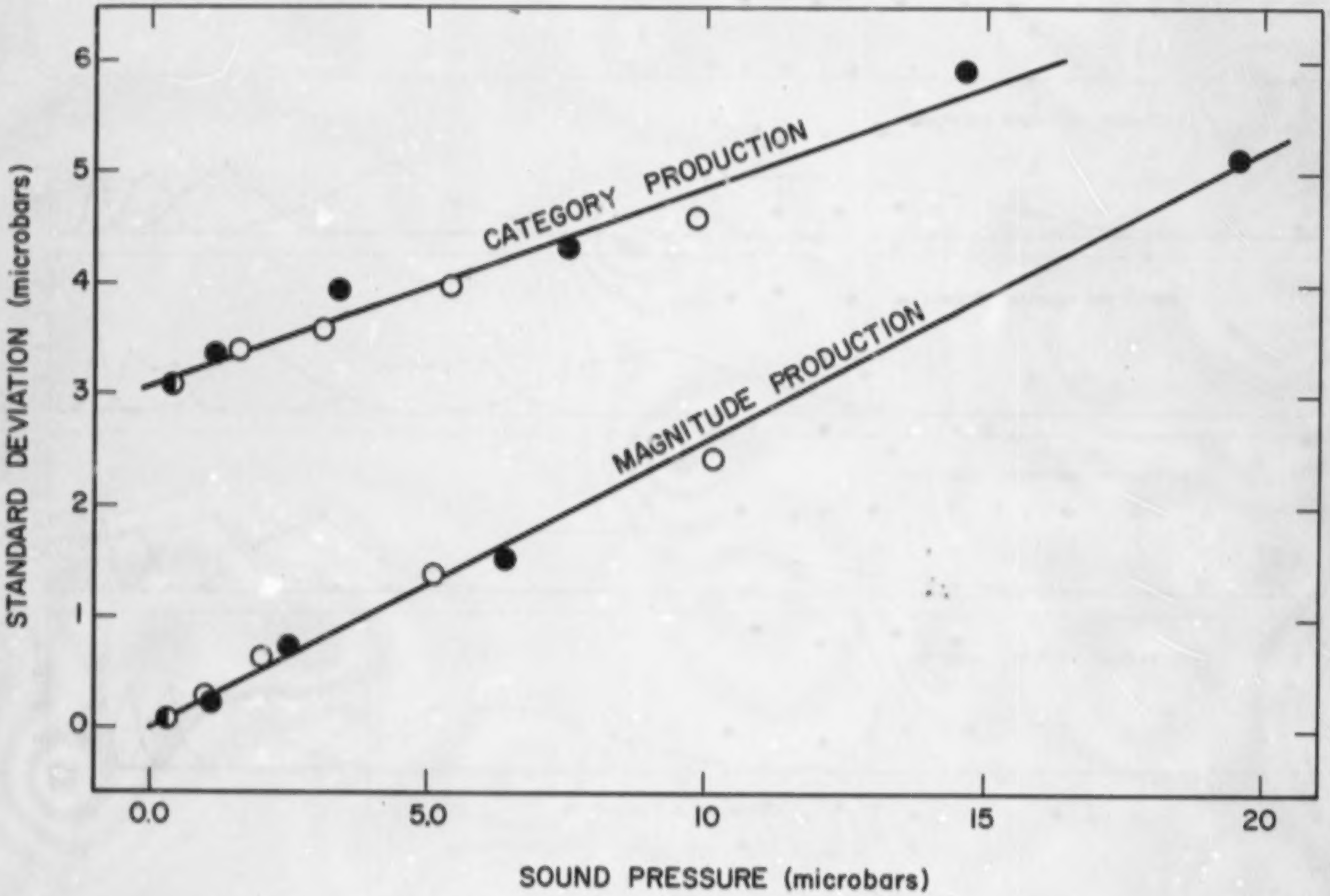

STANDARD DEVIATION (microbars)
0
1
2
3
4
5
6
0.0
5.0
10
15
20
SOUND PRESSURE (microbars)
CATEGORY PRODUCTION
MAGNITUDE PRODUCTION

Fig. 9

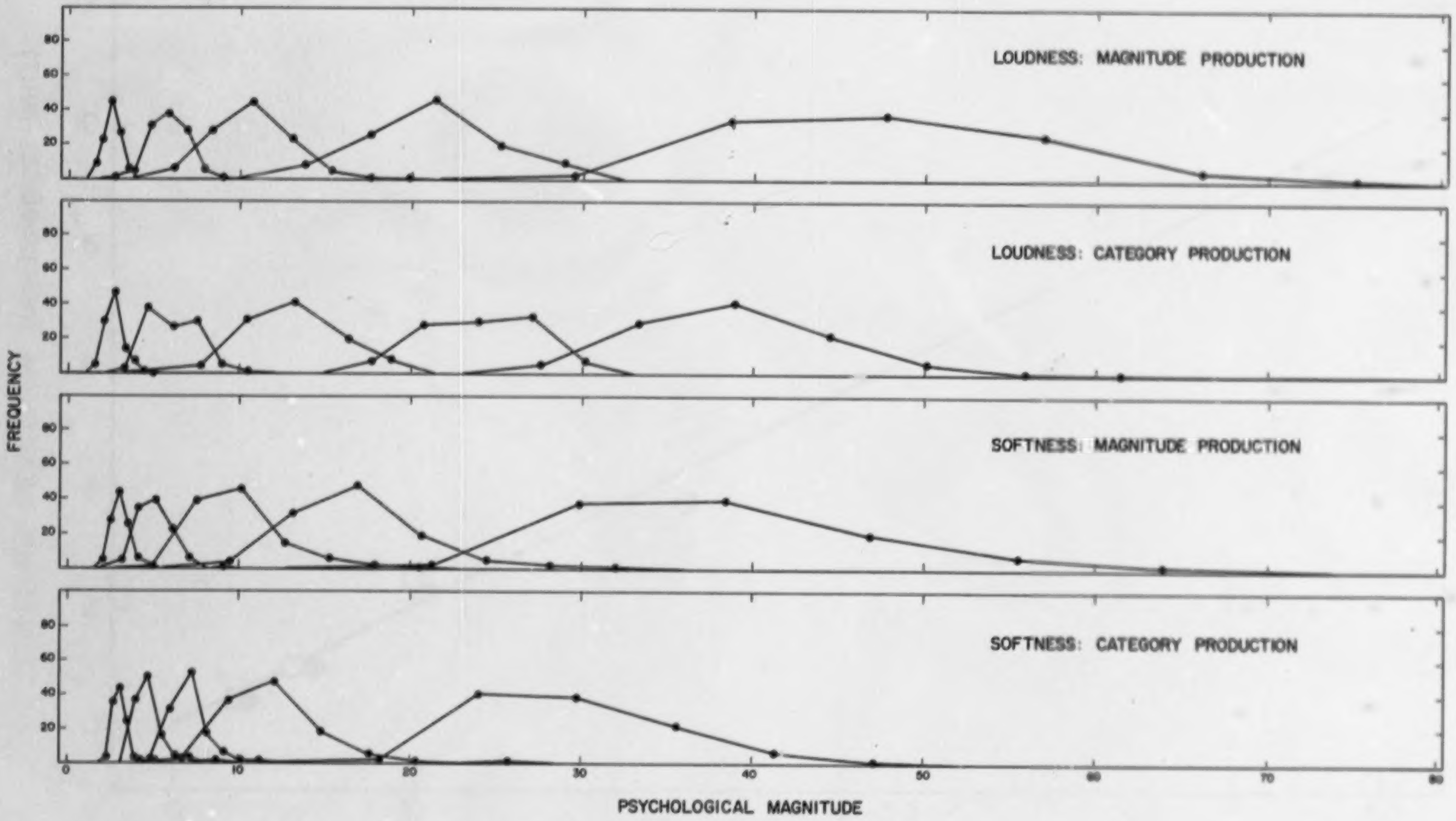

LOUDNESS: MAGNITUDE PRODUCTION
LOUDNESS: CATEGORY PRODUCTION
SOFTNESS: MAGNITUDE PRODUCTION
SOFTNESS: CATEGORY PRODUCTION
FREQUENCY
80
60
40
20
0
10
20
30
40
50
60
70
80
PSYCHOLOGICAL MAGNITUDE

Fig. 10

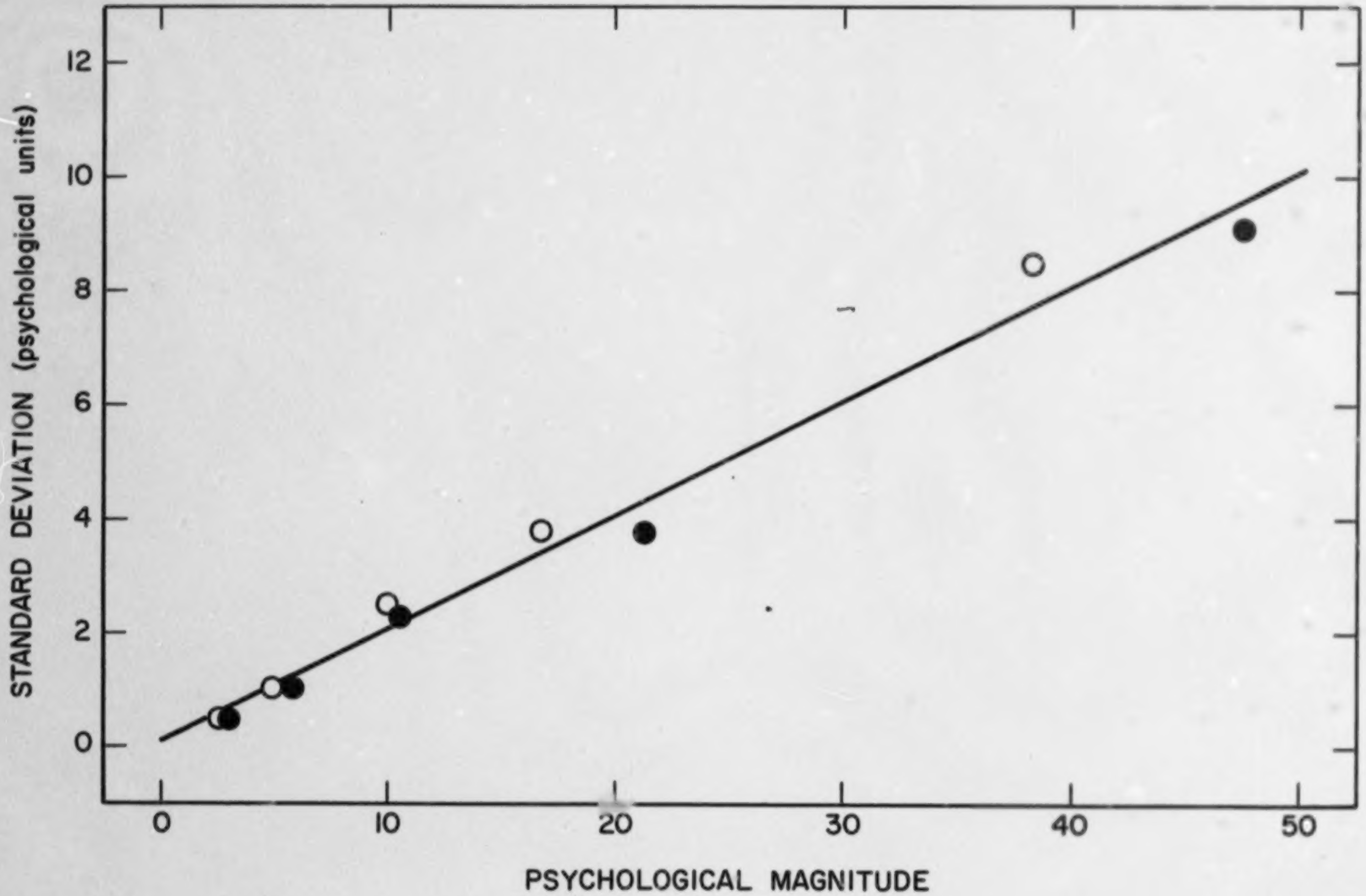

STANDARD DEVIATION (psychological units)
12
10
8
6
4
2
0
0
10
20
30
40
50
PSYCHOLOGICAL MAGNITUDE

Fig. 11

THE AUTOPHONIC SCALE OF VOICE LEVEL FOR CONGENITALLY DEAF SUBJECTS

Harlan Lane
The University of Michigan

The Autophonic Scale of Voice Level for Congenitally Deaf Subjects

Harlan Lane
Behavior Analysis Laboratory
University of Michigan

The autophonic scale of voice level relates the speaker's estimate of his own vocal level to the sound pressure level produced. The reception scale of voice level relates the listener's estimate of a speaker's loudness to the sound pressure level presented. Prior research, employing several psychophysical methods for determining and validating these subjective scales, has shown that the sensory dynamics of speaking and listening are quite different. The outcome of cross-modality matching, in which _S_s equate their vocal level to sounds presented at various intensities, is predicted by the disparate exponents of the power functions that describe the autophonic and reception scales, as is the outcome of cross-modality compensation, in which _S_s adjust their vocal level, under changes in the amplification of sidetone, to maintain sidetone loudness constant. Furthermore, increased or attenuated voice sidetone, or even total masking of sidetone, has no appreciable effect on the autophonic scale (Lane, Catania, & Stevens, 1961).

These findings and others (Lane, 1963) permit the inference that the speaker does not rely on loudness in judging autophonic level and the listener does not rely on the speaker's autophonic level in judging his loudness. The present study adds further support to this conclusion by showing that the autophonic scale of voice level for congenitally deaf speakers, who have never had auditory cues available while vocalizing, is the same as that scale for normal _S_s, vocalizing with and without masking of sidetone.

Method

Subjects

Four female and three male, adolescent Ss were employed;[1] each had profound binaural deafness since birth. In five cases, one or both parents was profoundly deaf, in a sixth the etiology was unknown, and in the seventh it was maternal rubella during pregnancy. The degree of hearing loss was determined by pure-tone audiometry (125-2,000 cps) and found always to exceed 70 db, and typically 100 db.

Apparatus and Procedure

The S sat at a distance of 18 in from a microphone (Altec, 633 A), connected to a tape recorder (Ampex 351). All instructions were given to an interpreter who then communicated with the S in sign language. First the S was asked to produce a loud vowel ("very big sound"), while the E set the recording level to zero VU. Five segments of a meterstick were then presented; the lengths of the segments were printed on their faces: 2.5, 5, 10, 20, and 40 in. The S was told to produce a vowel at moderate level and that the size of the sound he had just produced corresponded to the size of the middle segment ("10"). Other segments were then presented in irregular order, eight times each, and the speaker was asked to respond to each with a vocal production of proportionate magnitude.

Tape recordings were processed with a graphic level recorder (Bruel and Kjaer 2305) and the relative level of each vocal response was read (in db) from the calibrated chart.

Results

Figure 1 shows the autophonic scale of voice level for the seven deaf Ss and, for comparison,[2] that of 24 normal Ss vocalizing with sidetone and without (110 db masking noise). The growth of apparent vocal level as a function of the sound pressure produced is well described, for all three experimental conditions, by a straight line in log-log coordinates, in other words, a power function (Stevens, 1960). The slope (exponent) of this function, determined by a least-squares fit to the means of the decibel levels produced, is much the same in the three cases: 1.2 for normal Ss vocalizing with sidetone, 1.3 when sidetone is absent because of masking or deafness.

Table 1 permits comparison of the variability in vocal productions by the three groups of Ss. Because the method of magnitude production was employed, the reference vocal level ("10") varied from one speaker to the next. The component of the total variance that was due to the S's choice of modulus was removed by subtracting the mean of all the vocal responses by a given S from that mean for the entire group, and then adding the difference to each vocal production by that S. This operation left unchanged the slope of the autophonic scale for each S, but it minimized the sum of the squared deviations of his productions around the regression line for the group. The standard deviations of the productions so treated are shown in Table 1. When sidetone is absent, because of masking deafness, the variability of vocal productions is not appreciably different than when it is present.

Discussion

The finding that the form and parameters of the autophonic scale for deaf speakers are the same as that for normal Ss confirms, in particular, that autophonic judgments need not be based on loudness and, in general, that the sensory dynamics of speaking and listening are usefully distinguished.

Our earlier research strongly supports this conclusion. First of all, the autophonic scale and the reception scale for voice level have drastically different exponents, 1.1 and 0.4 respectively (Lane, 1963). This disparity is borne out quantitatively by experiments that require Ss to match their autophonic level to sounds of varying loudness and by complementary experiments that require Ss to decrease autophonic level, under increases in sidetone level, to maintain sidetone loudness constant. Both kinds of equal-sensation functions, which turn out to be reciprocal power functions, have an exponent predicted by the ratio of the exponents of the autophonic and reception scales for voice level (Lane, 1963; Lane, Catania, & Stevens, 1961).

Further evidence that autophonic judgments are not based on loudness has come from experiments in which the normal relations between autophonic level and sidetone level are altered radically. One-hundred-fold amplification or attenuation of sidetone level, or total masking of sidetone, leaves the autophonic scale essentially unchanged. Moreover, the speaker readily maintains a constant vocal level, when instructed to do so, despite changes in sidetone level as great as 80 db.

On the other side of the coin, it has been shown that loudness estimates are not based on autophonic level. Magnetic recording techniques permit autophonic levels to be decoupled from their corresponding sound pressure levels so that the two may be varied independently and the contribution of each to speech-loudness estimates may be assessed. When autophonic level is varied from whispering at the one extreme to shouting at the other, the quality (spectrum) of the voice inevitably varies, but the reception scale for voice level remains invariant (Lane, 1963).

Although the conclusion seems inescapable that estimates of loudness are quite different from estimates of autophonic level, Warren (1962) has recently contended, on the contrary, that they are the same, and that both kinds of estimates are based on spatial localization. Observers learn to judge the loudness of externally produced sounds, he suggests, by noting how their loudness changes with distance. Self generated sounds do not, of course, vary in distance, but the speaker has learned, so the theory goes, how much he must increase his loudness to reach listeners at varying distances.

The "physical-correlate theory" (Warren, 1958) has been criticised previously (Stevens, 1959). This more recent extension of the theory, relating distance, loudness and autophonic level, seems faulty in all three correspondences. First, the prediction that loudness grows linearly with sound pressure, which follows from the inverse-square law relating distance to sound intensity, is counter to the findings of innumerable studies.[3] Second, the contention that estimates of autophonic level and of loudness are identical is counter to the several findings cited above. Warren disputes the conclusiveness of one of these findings, namely that the autophonic scale is invariant under total masking of sidetone. He suggests that the speaker's familiarity with the relation between autophonic and sidetone level provides a basis for loudness judgments even in the absence of auditory cues. This explanation of the invariance of the autophonic scale cannot apply, however, to the congenitally deaf Ss who, lacking this familiarity, nevertheless provided the same subjective scale as did normal speakers.

Finally, the hypothesis that a speaker's estimate of his autophonic level is in fact an estimate of his distance from a listener, on those prior occasions when he spoke at that loudness, notably lacks support in the present confirmation of the autophonic scale with congenitally deaf Ss.

The sensory basis for autophonic judgments is more likely to be found among the cues to vocal effort. Ladefoged and McKinney (1963) report that the rate of work done upon the air in producing a voiced sound is proportional to the 1.3 power of the sound pressure. Since the autophonic scale has the same form and exponent as this vocal work function, the authors suggest that estimates of autophonic level are based on physiological effort — which is what our Ss have been telling us for some years now.

Summary

The autophonic scale of voice level was determined for seven Ss with profound congenital deafness and found to be the same as that for speakers with normal hearing. Several kinds of evidence support the conclusion that the speaker does not rely on loudness in judging autophonic level and the listener does not rely on the speaker's autophonic level in judging his loudness.

References

Ladefoged, P. & McKinney, N. P. Vocal effort, intensity, and loudness. J. Acoust. Soc. Amer., 1963 (in press).

Lane, H. L. Psychophysical parameters of vowel perception. Psychol. Monogr., 1963 (in press).

Lane, H. L., Catania, A. C., & Stevens, S. S. Voice level: autophonic scale, perceived loudness, and effects of sidetone. J. Acoust. Soc. Amer., 1961, 33, 160-167.

Stevens, S. S. On the validity of the loudness scale. J. Acoust. Soc. Amer., 1959, 31, 995-1003.

Stevens, S. S. The psychophysics of sensory function. Amer. Scient., 1960, 48, 226-253.

Stevens, S. S. & Guirao, M. Loudness, reciprocality, and partition scales. J. Acoust. Soc. Amer., 1962, 34, 1466-1471.

Warren, R. M. Are 'autophonic' judgments based on loudness? Amer. J. Psychol., 1962, 75, 452-456.

Warren, R. M. A basis for loudness judgments. Amer. J. Psychol., 1958, 71, 700-709.

Footnotes

1. The cooperation of the Michigan School for the Deaf, Dr. B. R. Siders, Superintendent, is gratefully acknowledged.

2. From Lane, Catania, and Stevens (1961).

3. The most recent study is by Stevens and Guirao (1962).

Table 1

The standard deviation of vocal levels (in db) in a determination of the autophonic scale by the method of magnitude production. (The number of subjects is denoted by *S*, the number of determinations at each criterion value by *N*.)

Criterion Value	Normal Ss (sidetone)	Normal Ss (no sidetone)	Deaf Ss
2.5	3.4	3.6	4.5
5	3.3	3.4	3.2
10	2.6	3.4	2.8
20	2.8	3.0	3.4
30	4.0	3.9	- -
40	- -	- -	4.5
S	24	24	7
N	72	72	56

Figure Caption

Fig. 1. The autophonic scale of voice level for seven speakers with profound congenital deafness, and for 24 Ss with normal hearing who vocalized with sidetone and without (110 db masking noise). Each point is the mean decibel level of 56 vocal productions by the deaf speakers and 72 by the normal Ss.

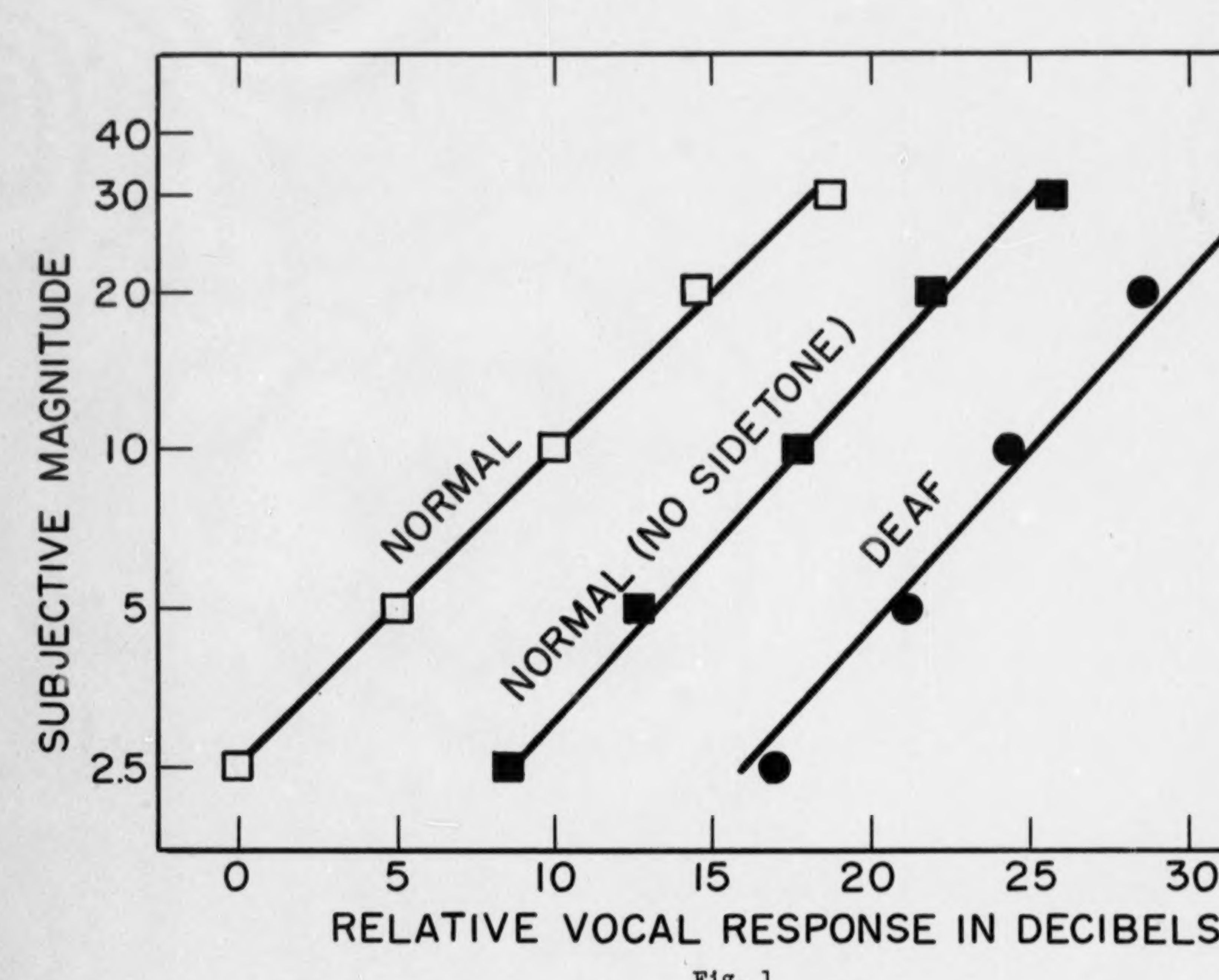

SUBJECTIVE MAGNITUDE
40
30
20
10
5
2.5
0
5
10
15
20
25
30
35
RELATIVE VOCAL RESPONSE IN DECIBELS
NORMAL
NORMAL (NO SIDETONE)
DEAF

Fig. 1

A NEW DETERMINATION OF THE PURE TONE EQUAL-LOUDNESS CONTOURS FOR EARPHONE LISTENING

James Kopp and Harlan Lane
The University of Michigan

A New Determination of the Pure Tone Equal-loudness Contours for Earphone Listening

James Kopp and Harlan Lane
Behavior Analysis Laboratory
University of Michigan

The observation that the loudness of a pure tone varies irregularly as a function of its frequency first received experimental study around the turn of the century. Fletcher and Wegel (1922) summarize early research on this function and describe its form for the special case in which the tone is at threshold intensity. With instrumentation that was crude by modern standards, these investigators concluded that the loudness of a pure tone decreases in a systematic fashion at frequencies below 1000 cps, and that, on the average, 70 times more intensity is required at 100 cps than at 1000 cps to yield a report of "just audible" (vide Fig. 1). The authors originally fitted a curve to these data which showed threshold intensity constant at frequencies over 1000 cps, but in a later study (Wegel, 1922) their data was fitted with a curve that sloped upward at high frequencies, in order to indicate the hypothetical intersection of the threshold curve and the curve of "maximum audibility" (threshold of feeling).

Kingsbury (1927) made the first determinations of loudness as a function of frequency at intensities above threshold. He presented pure tones monaurally through a high-impedance telephone receiver to 22 Ss. The Ss adjusted each tone to sound equal in loudness to a 700-cps tone at a predetermined intensity. "Contours" of equal loudness were fitted to the averaged data (Fig. 2) so as to yield a family of curves, parallel to the original threshold curve of Fletcher and Wegel.

A more comprehensive study of equal loudness relations was done by Fletcher and Munson (1933). Pure tones were presented binaurally through earphones to 11 Ss. The method of constant stimuli was employed with a standard set at 1000 cps. Measurements were made at 13 intensity levels over the range 0 to 120 phons. Equal loudness was taken to be the intensity of the comparison tone at a given frequency needed to elicit judgments either "greater" or "less" 50% of the time. The earphones were calibrated in a pilot study in which observers adjusted the loudness of a tone in the earphones to equal that of a free-field tone of known intensity. Thus the equal-loudness contours could be determined at known sound pressure levels. These contours are presented in Fig. 3a. They are irregular and change slope at several points in the frequency region above 1000 cps. Fletcher and Munson suggest that the discontinuities in the curves are due to diffraction of the sound around the head of the observers during the free-field calibration of the telephone receivers. The authors smoothed the contours for this reason and plotted them in decibels above the threshold curve they obtained separately (Fig. 3b).

Two subsequent studies (Churcher & King, 1937; Robinson & Dadson, 1956) have yielded pure tone equal-loudness contours (Figs. 4, 5) obtained by the method of constant stimuli in free-field listening conditions. These differ only slightly from the unsmoothed curves of the Fletcher and Munson study. A recent article by Lochner and Burger (1962) summarizes the latter three studies and develops a standard set of free-field equal-loudness contours (Fig. 6) and a description of their mathematical form. Equal-loudness contours have also been obtained under free-field conditions for bands of noise. The marked differences between these contours, and those for pure tones, especially at high frequencies, are discussed by Stevens (1956).

Most investigators distinguish between two families of equal-loudness contours, one for earphone and the other for free-field listening, although the former have received little study. Stevens and Davis (1938) present the data reported by Fletcher and Munson (Fig. 3a) and then replot these data relative to a threshold function for earphone listening (Sivian and White, 1933, Fig. 7) in order to estimate equal-loudness contours under these conditions. Beranek (1949) reprints these two forms of the Fletcher and Munson equal-loudness data with some additions; he represents the original plot as obtained under free-field conditions (which is not strictly true) and the second set of curves as obtained independently under closed-ear conditions (also not correct). Although there are _a priori_ grounds for expecting differences between equal-loudness contours obtained under the two conditions (Beranek, 1949, p. 731), the limited data available do not provide a basis for the distinction. Zwicker and Feldtkeller (1955) are the only investigators since Kingsbury who, to our knowledge, have measured these contours with earphone listening. Using a continuous tracking method, they obtained contours that were quite similar in form to those determined under free-field conditions (Fig. 8).

Two limitations that may be ascribed to most of the prior research on equal-loudness contours serve as the point-of-departure for the present study: first, there is a dearth of research employing earphone listening, although data obtained under these conditions should prove valuable both for psycho-acoustic theory and applications such as magnetic reproduction of sound; second, the psychophysical methods employed heretofore do not permit a detailed description of local changes in the equal-loudness contours.

Method

Pure tones were generated by a beat frequency oscillator (Bruel and Kjaer, 1014) which slowly swept the audio-frequency spectrum. The oscillator supplied two earphones (Grason-Stadler TDH-39), matched in their frequency characteristics. One earphone was mounted without cushion on a binaural headset along with a "dummy" earphone, and the other was similarly mounted on an artificial ear (Bruel and Kjaer, 4151) housed in a sound-attenuating box. The artificial ear contained a 6-cc coupler (ASA Type 1) and a condenser microphone (Bruel and Kjaer, 4132) whose output was amplified (Bruel and Kjaer, 2603) and sent to a compressor incorporated in the oscillator. The compressor signal compensated the oscillator output for the irregular frequency-response characteristic of the earphones, so that a constant sound pressure (±1 db) was provided at the Ss' earphone as the signal frequency changed from 50-5000 cps.

In order to determine equal-loudness contours, the audible spectrum was swept from 1000 cps upward to 5000 cps and then from 1000 cps downward to 50 cps at a rate of 4 octaves/min. The S adjusted a sone potentiometer (Poulton & Stevens, 1955) so as to maintain loudness constant. Changes in the potentiometer setting during the frequency sweep were recorded by sending a constant voltage through a second sone potentiometer which was "ganged" to the first, and then recording the output on a calibrated graphic level recorder (Bruel and Kjaer, 2305), mechanically synchronized with the sweep of the oscillator. Hence, the change in sound pressure level required to maintain equal loudness at a given frequency was recorded (in decibels) directly over that frequency. This method yields a continuous plot of equal loudness at any preset phon level, for frequencies from 50 to 5000 cps. A schematic of the apparatus is shown in Fig. 9. A graph of the frequency response of the earphone, the compensatory output of the compressor, and the frequency response of the earphone when in the compressor circuit are shown in Fig. 10.

Initial data were obtained from two untrained, male observers, 19 years of age, with normal hearing. Equal-loudness contours were determined for one S at 40 and 80 phons and for the other at 60 and 80 phons. From four to seven replications were made for each S at each intensity level.

Results and Discussion

Figures 11 and 12 are photographic overlays of the equal loudness contours for the two Ss. The 40-phon contour of S (1) and both contours of S (2) flatten out below approximately 300 cps because the sone potentiometer used in this initial study had insufficient range. The deviant high-frequency contour shown for S (1) was obtained with the earphone mounted in a neoprene cushion.

Although the contours differ in form for the two Ss, the replications for any one S are quite reliable. The tracking method for determining equal-loudness contours seems to constitute a convenient and reliable psychophysical technique. Figure 13 permits comparison of the equal-loudness relations obtained in our initial study with those of Fletcher and Munson. The figure shows the effect of intensity of the standard tone on the form of the equal-loudness contour by plotting the difference (in db) between corresponding points on the contours at two intensity levels. The functions obtained in the two studies are similar in this respect. The present research will be extended to examine more _S_s and intensities and the effects of certain parameters such as the frequency and recurrence of the reference tone.

References

Beranek, L. L. *Acoustic Measurement*. New York: John Wiley & Sons, Inc., 1949, 200-201.

Churcher, B. G. & King, A. J. The performance of noise meters in terms of the primary standard. *J. Instn. Electrical Engrs.*, 1937, 81, 57-90.

Fletcher, H. & Munson, W. A. Loudness, its definition, measurement, and calculation. *J. Acoust. Soc. Amer.*, 1933, 5, 82-108.

Fletcher, H. & Wegel, R. L. The frequency sensitivity of normal ears. *Phys. Rev.*, 1922, 19, 553-565.

Kingsbury, B. A. A direct comparison of the loudness of pure tones. *Phys. Rev.*, 1927, 29, 588-600.

Lochner, J. P. A. & Burger, J. F. Pure-tone loudness relations. *J. Acoust. Soc. Amer.*, 1962, 34, 576-581.

Robinson, D. W. & Dadson, R. S. A redetermination of the equal-loudness relations for pure tones. *Brit. J. Appl. Phys.*, 1956, 7, 166-181.

Sivian, L. J. & White, S. D. On minimum audible sound fields. *J. Acoust. Soc. Amer.*, 1933, 4, 288-321.

Stevens, S. S. Calculation of the loudness of complex noise. *J. Acoust. Soc. Amer.*, 1956, 28, 807-832.

Stevens, S. S. & Davis, H. *Hearing: its psychology and physiology*. New York: John Wiley & Sons, Inc., 1938, 124.

Stevens, S. S. & Poulton, E. C. On the halving and doubling of the loudness of white noise. *J. exp. Psychol.*, 1956, 51, 71-78.

Wegel, R. L. The physical characteristics of audition and dynamical analysis of the external ear. *Bell Syst. Tech. J.*, 1922, 1, 56-58.

Zwicker, E. & Feldtkeller, T. Über die lautstärke von gleichförmigen geräuschen. *Acustica*, 1955, 5, 303-316.

Figure Captions

Fig. 1. Early findings on the threshold intensity of pure tones as a function of their frequency. After Fletcher & Wegel (1922). Sound sources included tuning forks (Wead), whistles (Rayleigh) and a vacuum tube oscillator powering a standard telephone receiver (Fletcher & Wegel).

Fig. 2. Equal loudness contours for earphone listening from Kingsbury (1927). Ordinate values are in db re: 1 dyne/cm^2. (-74 db re: 1 dyne/cm^2 equals 0 db re: .0002 dyne/cm^2.)

Fig. 3. Alternate plots of equal loudness relations from Fletcher & Munson (1933). Earphones were calibrated by separate equal loudness matches in a free-field. a) Curves fitted to the original data plotted in db re: .0002 dyne/cm^2, b) Smoothed curves plotted in db above the obtained threshold curve.

Fig. 4. Contours of equal loudness obtained under free-field conditions. The 110-phon curve was constructed by extrapolation. From Churcher & King (1937).

Fig. 5. Free-field equal loudness contours from Robinson & Dadson (1956), showing high frequency differences due to age. Dotted curves are for subjects whose average age is 60 years. Solid curves are for subjects averaging 20 years. Threshold curve (MAF) is from Sivian & White (1933).

Fig. 6. Generalized free-field equal-loudness contours from Lochner & Burger (1962). The threshold curve (MAF) is from Sivian & White (1933).

Fig. 7. The auditory threshold as a function of frequency under three conditions of listening. From Sivian & White (1933). Ordinate in db re: 1 dyne/cm^2. (-74 db re: 1 dyne/cm^2 equals 0 db re: .0002 dyne/cm^2.) Curve 1 represents a best fit to data obtained in nine different studies over a ten-year period. Curves 2 and 3 are fitted to data obtained by Sivian & White.

Fig. 8. Equal loudness contours obtained by a continuous tracking method for one subject. From Zwicker & Feldtkeller (1955). The curves were constructed by sampling average values through the tracks. Ordinate values are db re: .0002 dynes/cm^2.

Fig. 9. Schematic diagram of apparatus used in the present study of pure-tone equal-loudness contours (see text).

Fig. 10. Frequency-response functions illustrating the method of compensating for a non-linear earphone response from 50 to 5000 cps. Curve A is the normal applied voltage response of the earphone (Grason-Stadler TDH-39) when supplied with a constant voltage from a beat frequency oscillator. Curve C is the compensated output of the beat frequency oscillator due to negative feedback from a matched earphone mounted on an artificial ear. Curve B is the resultant sound pressure from the original earphone.

Fig. 11. Continuous equal-loudness contours for subject 1. The reference frequency is 1000 cps. The deviant contours between 2000 and 5000 cps were obtained with doughnut cushion on the earphone.

Fig. 12. Continuous equal-loudness contours for subject 2.

Fig. 13. Effect of intensity level on the spacing of equal-loudness contours determined by Fletcher & Munson (solid curves) and in this study (dashed curves). The difference (in db) between corresponding points on a pair of equal-loudness contours is shown as a function of their frequency.

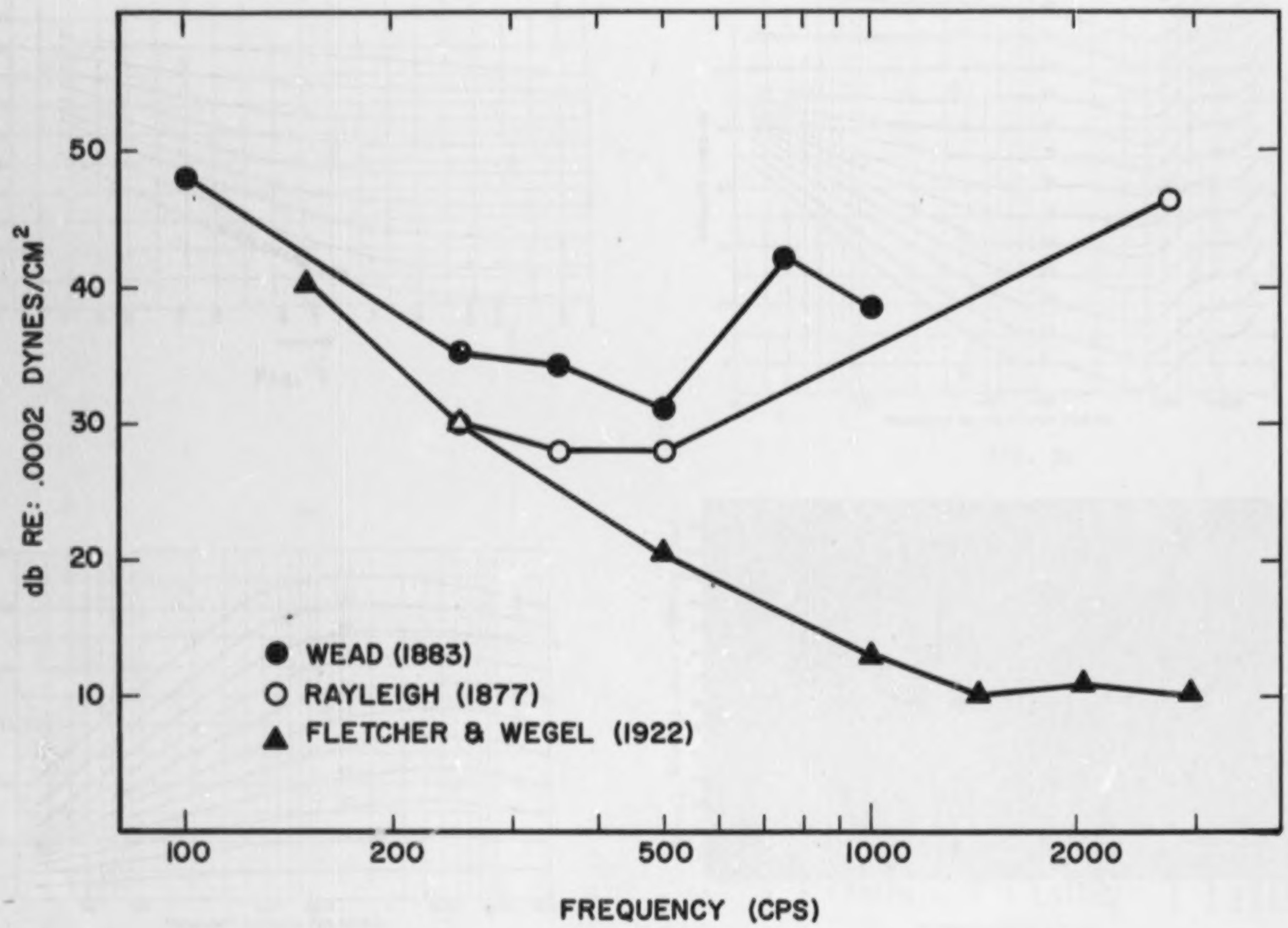

50
40
30
20
10
db RE: .0002 DYNES/CM²
100
200
500
1000
2000
FREQUENCY (CPS)
● WEAD (1883)
○ RAYLEIGH (1877)
▲ FLETCHER & WEGEL (1922)

Fig. 1

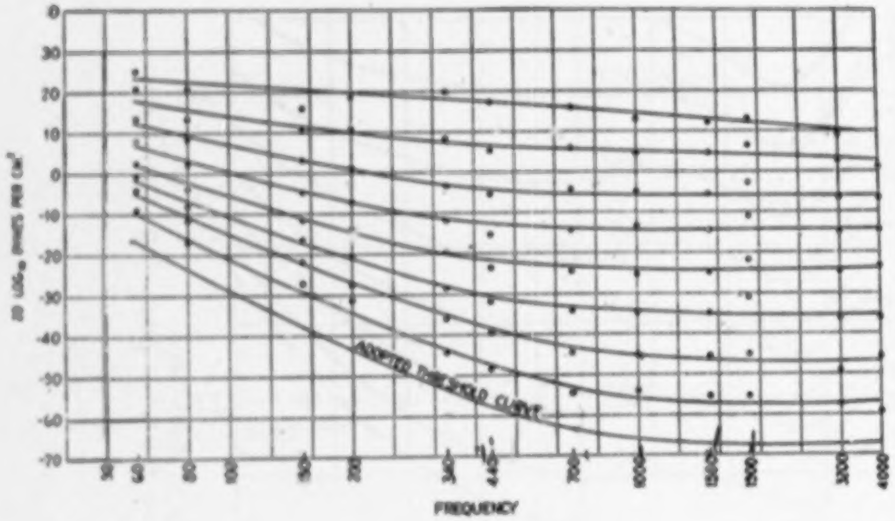


Fig. 2

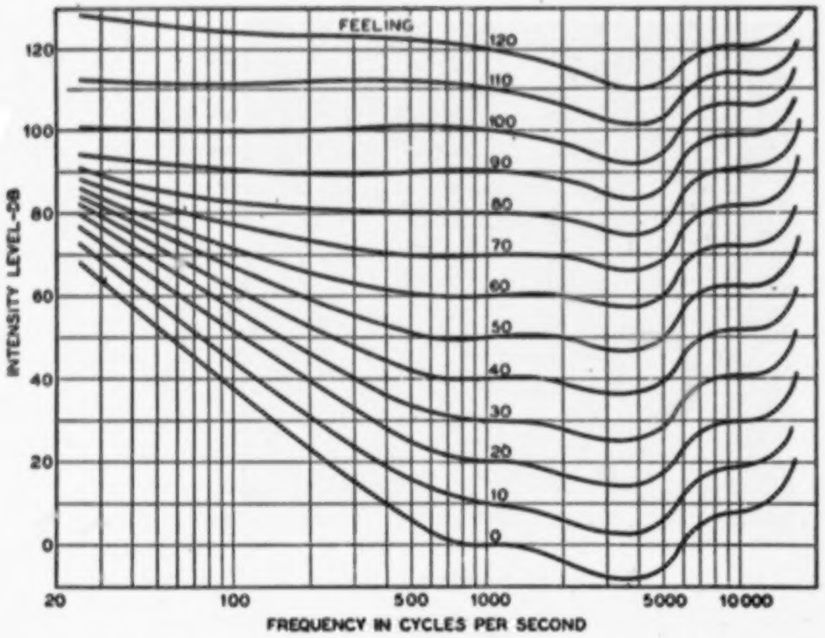


Fig. 3a

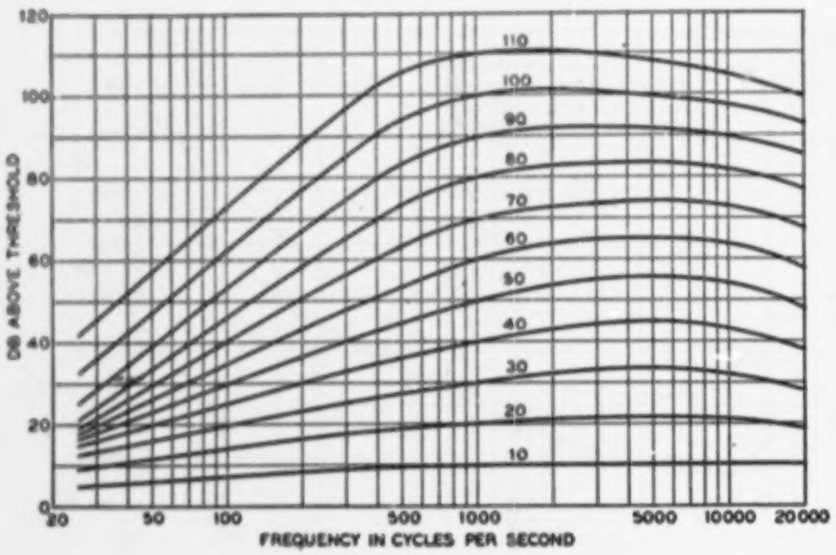


Fig. 3b

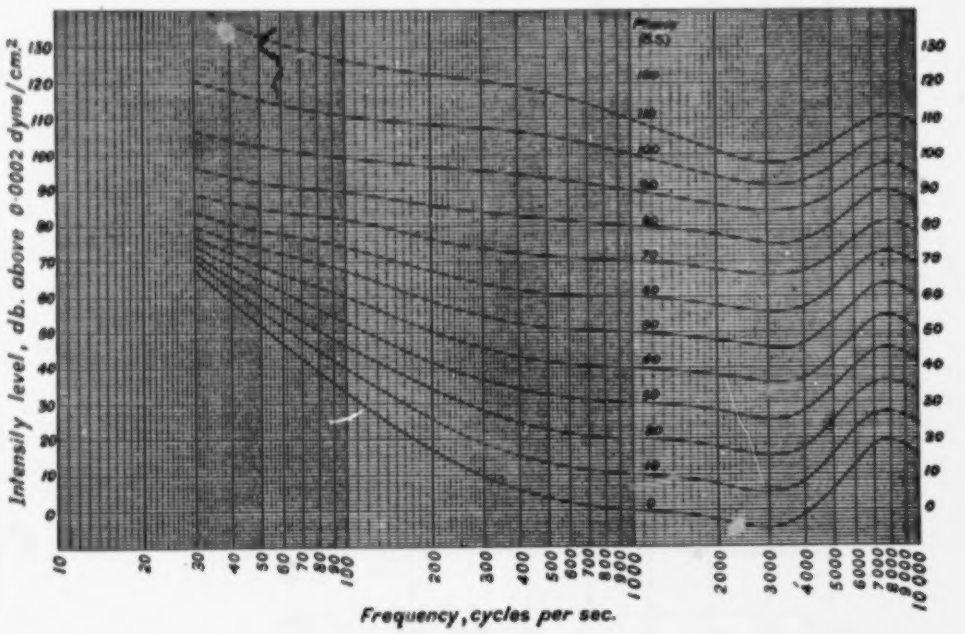


Fig. 4

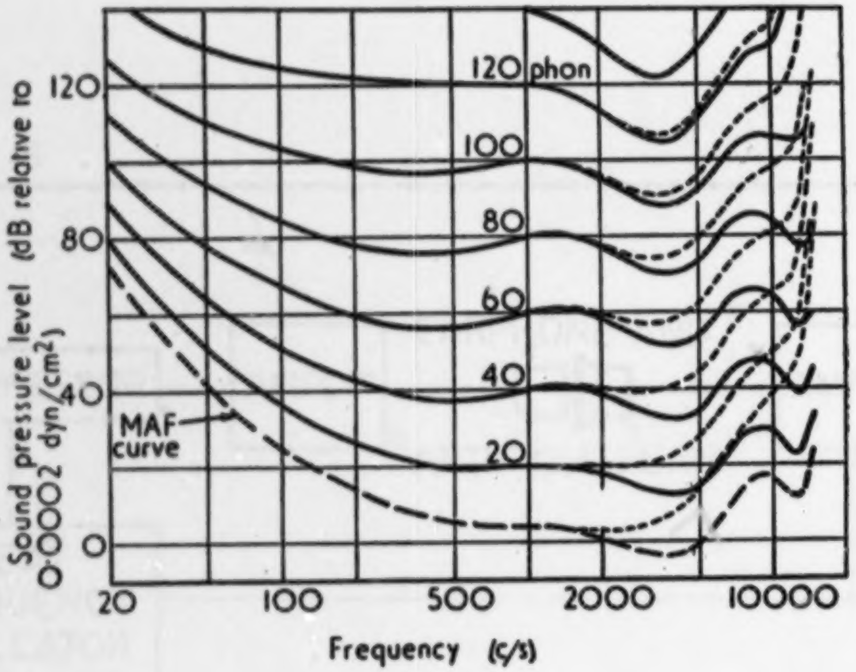


Fig. 5

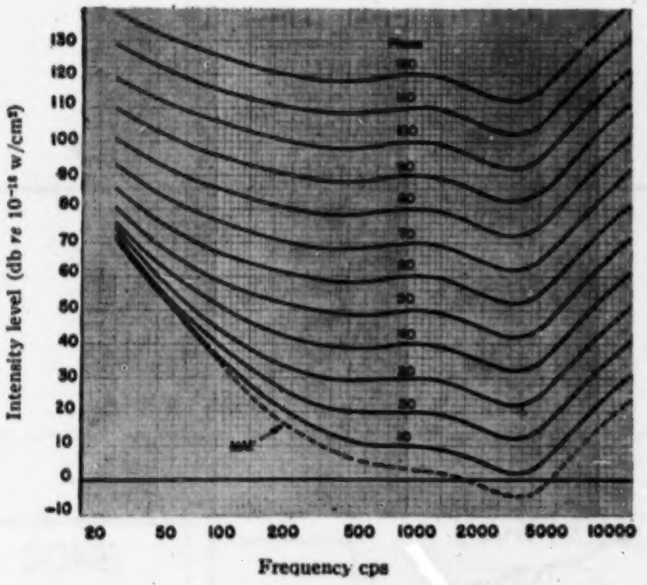


Fig. 6

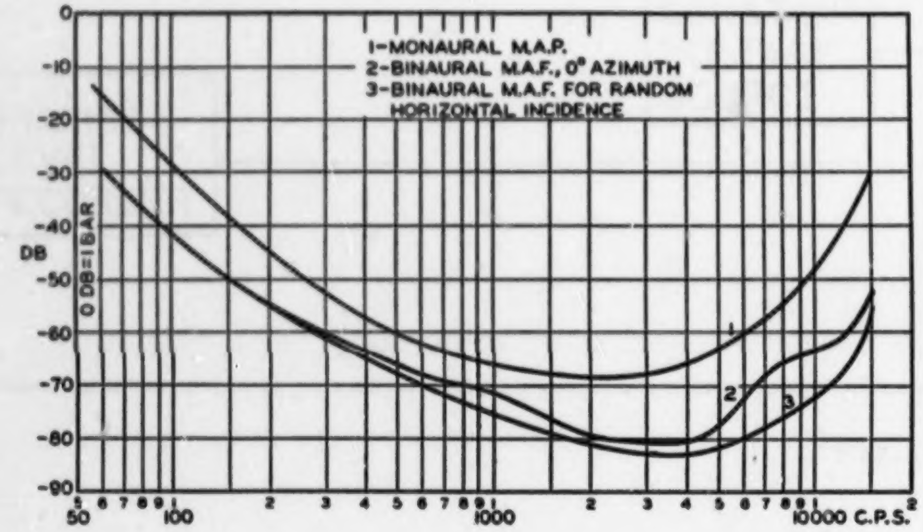


Fig. 7

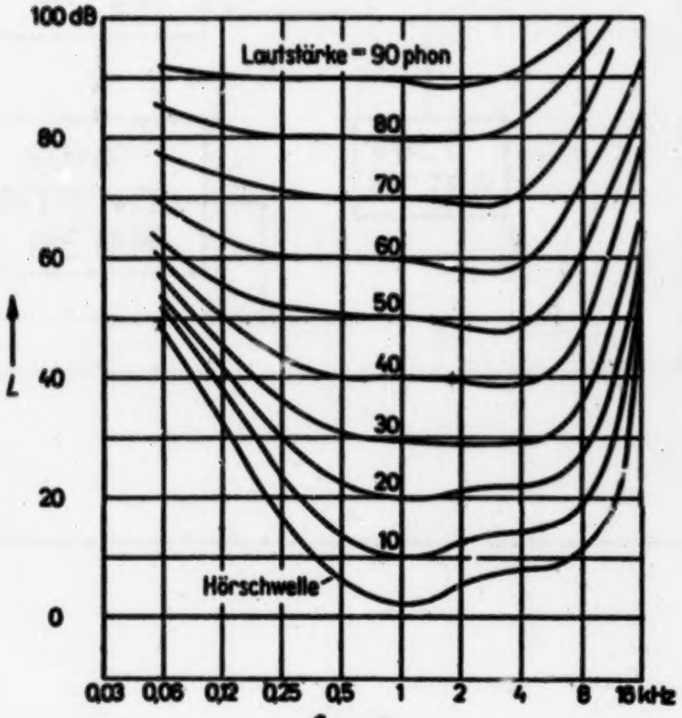


Fig. 8

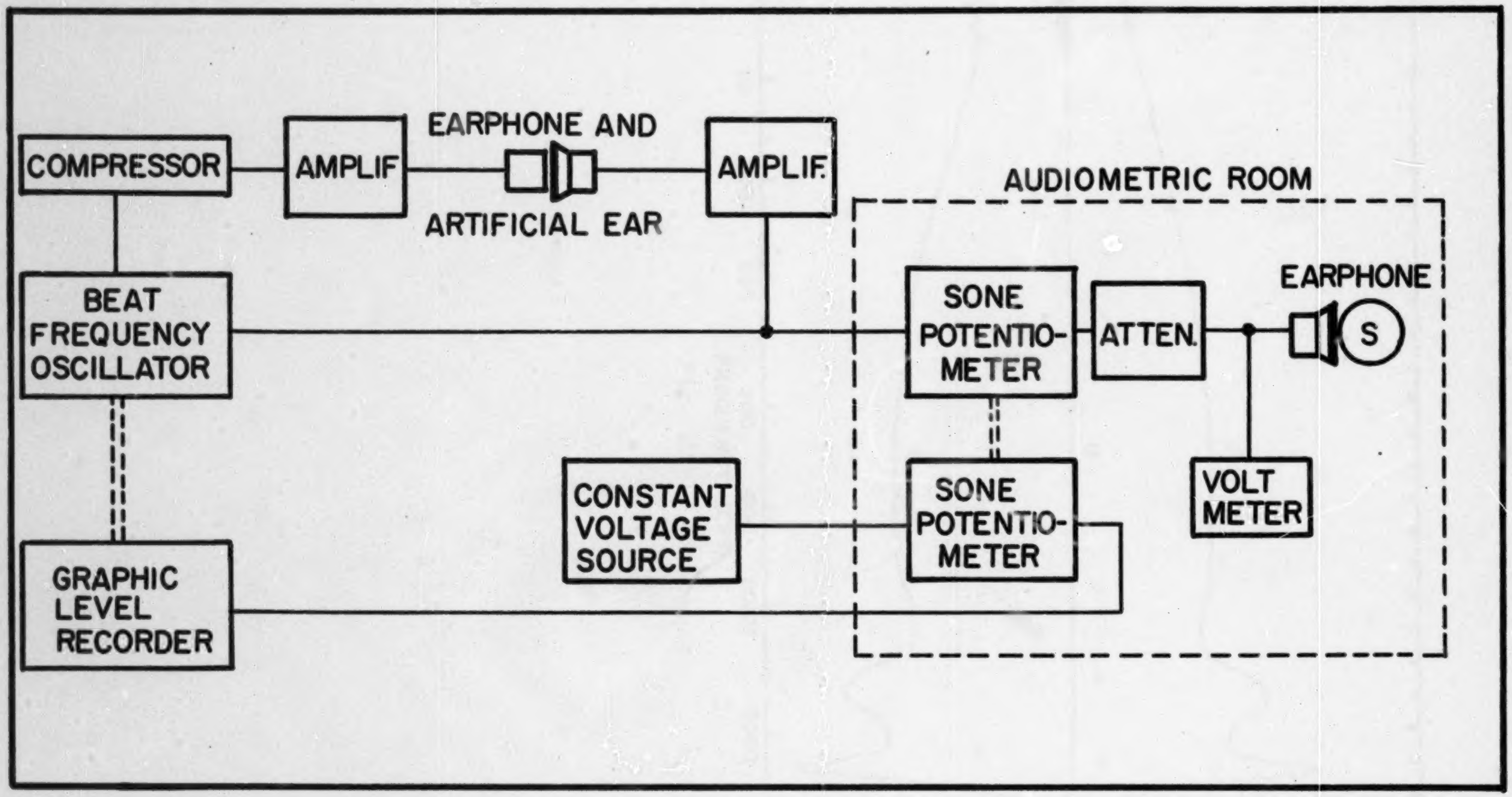

COMPRESSOR
AMPLIF
EARPHONE AND
ARTIFICIAL EAR
AMPLIF.
AUDIOMETRIC ROOM
BEAT FREQUENCY OSCILLATOR
SONE POTENTIO-METER
ATTEN.
EARPHONE
S
CONSTANT VOLTAGE SOURCE
SONE POTENTIO-METER
VOLT METER
GRAPHIC LEVEL RECORDER

Fig. 9

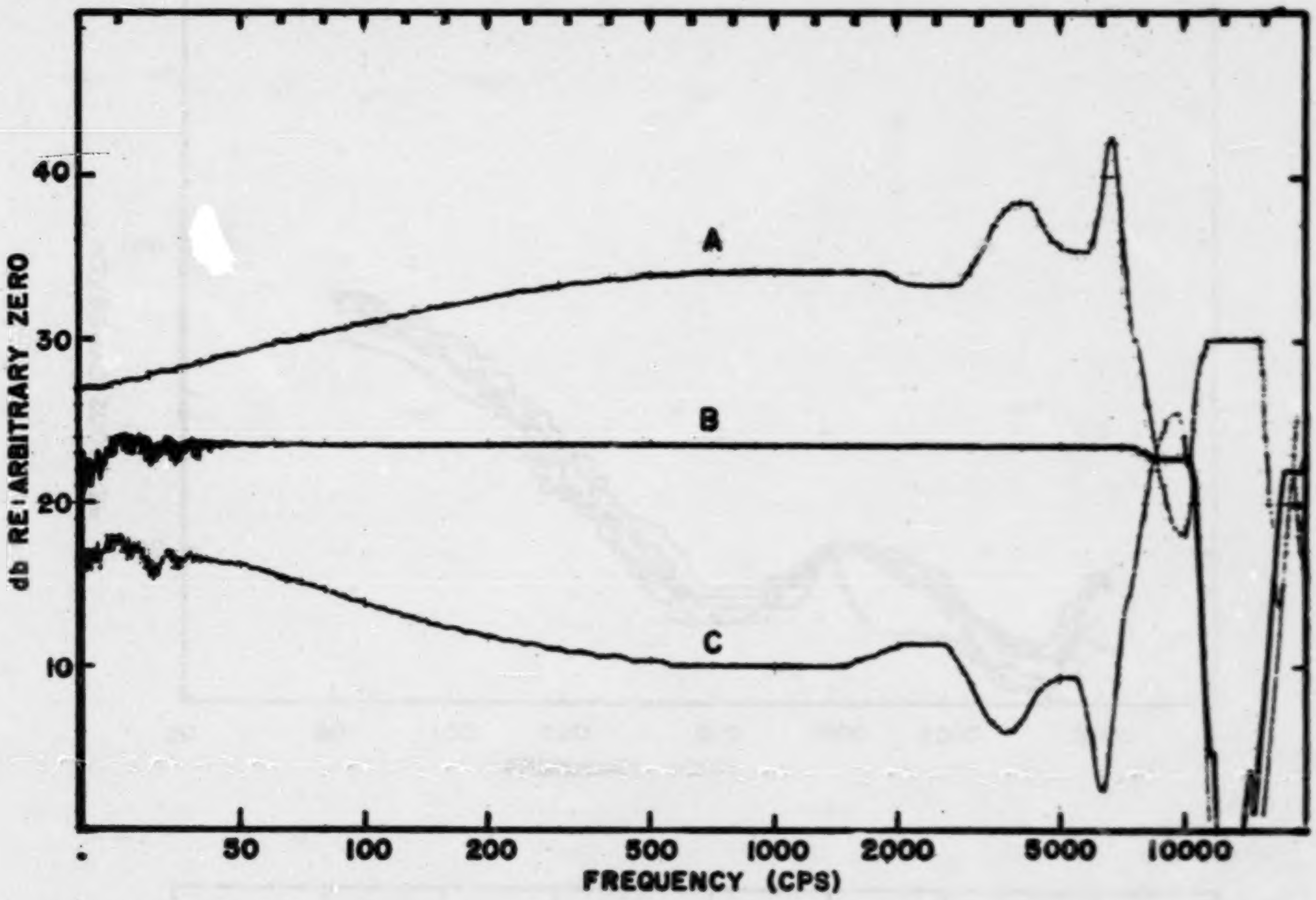

A
B
C
40
30
20
10
db RE: ARBITRARY ZERO
50
100
200
500
1000
2000
5000
10000
FREQUENCY (CPS)

Fig. 10

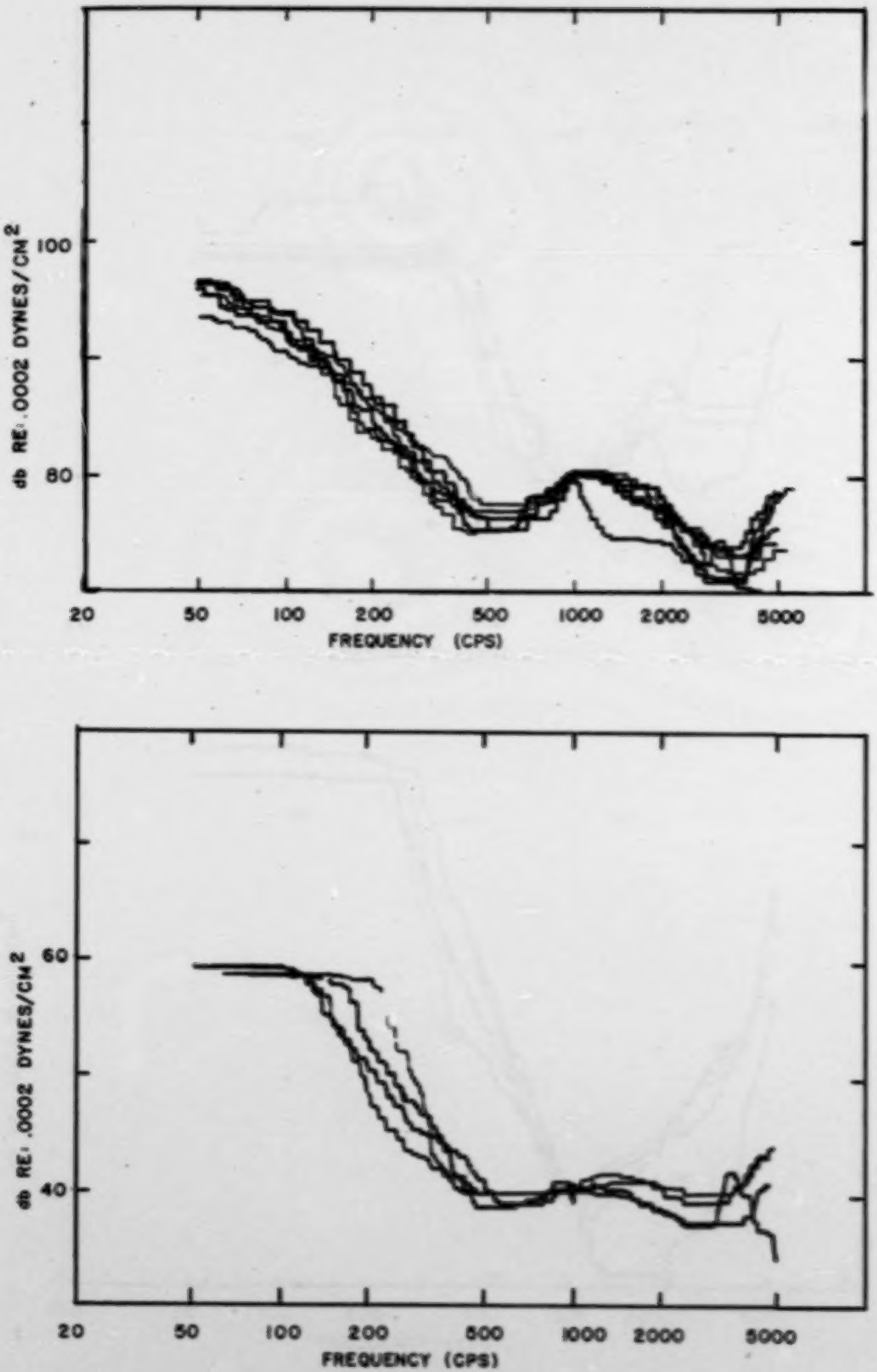

db RE: .0002 DYNES/CM²
100
80
20
50
100
200
500
1000
2000
5000
FREQUENCY (CPS)
db RE: .0002 DYNES/CM²
60
40
20
50
100
200
500
1000
2000
5000
FREQUENCY (CPS)

Fig. 11

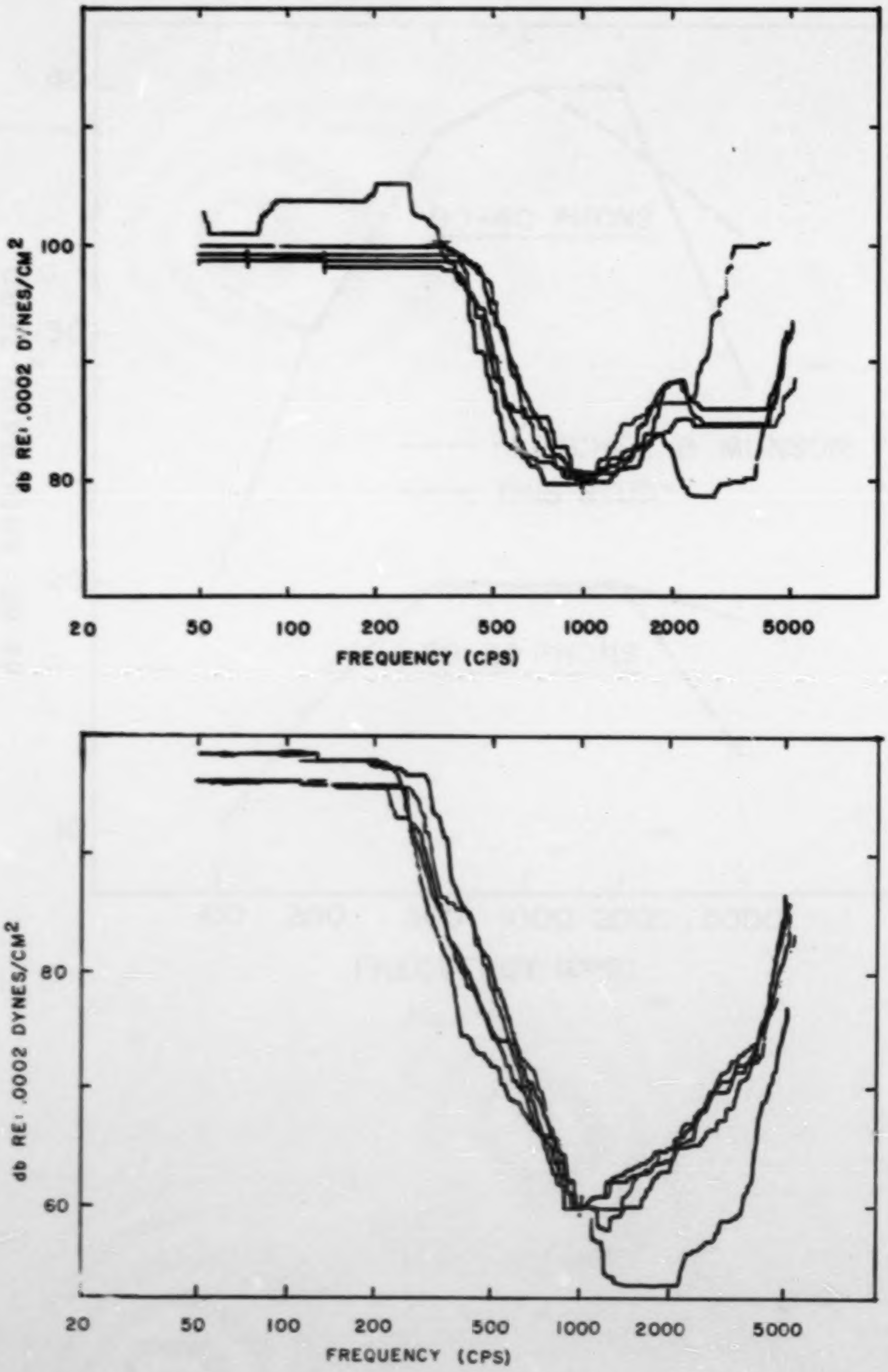

db RE: .0002 DYNES/CM²
100
80
20
50
100
200
500
1000
2000
5000
FREQUENCY (CPS)
db RE: .0002 DYNES/CM²
80
60
20
50
100
200
500
1000
2000
5000
FREQUENCY (CPS)

Fig. 12

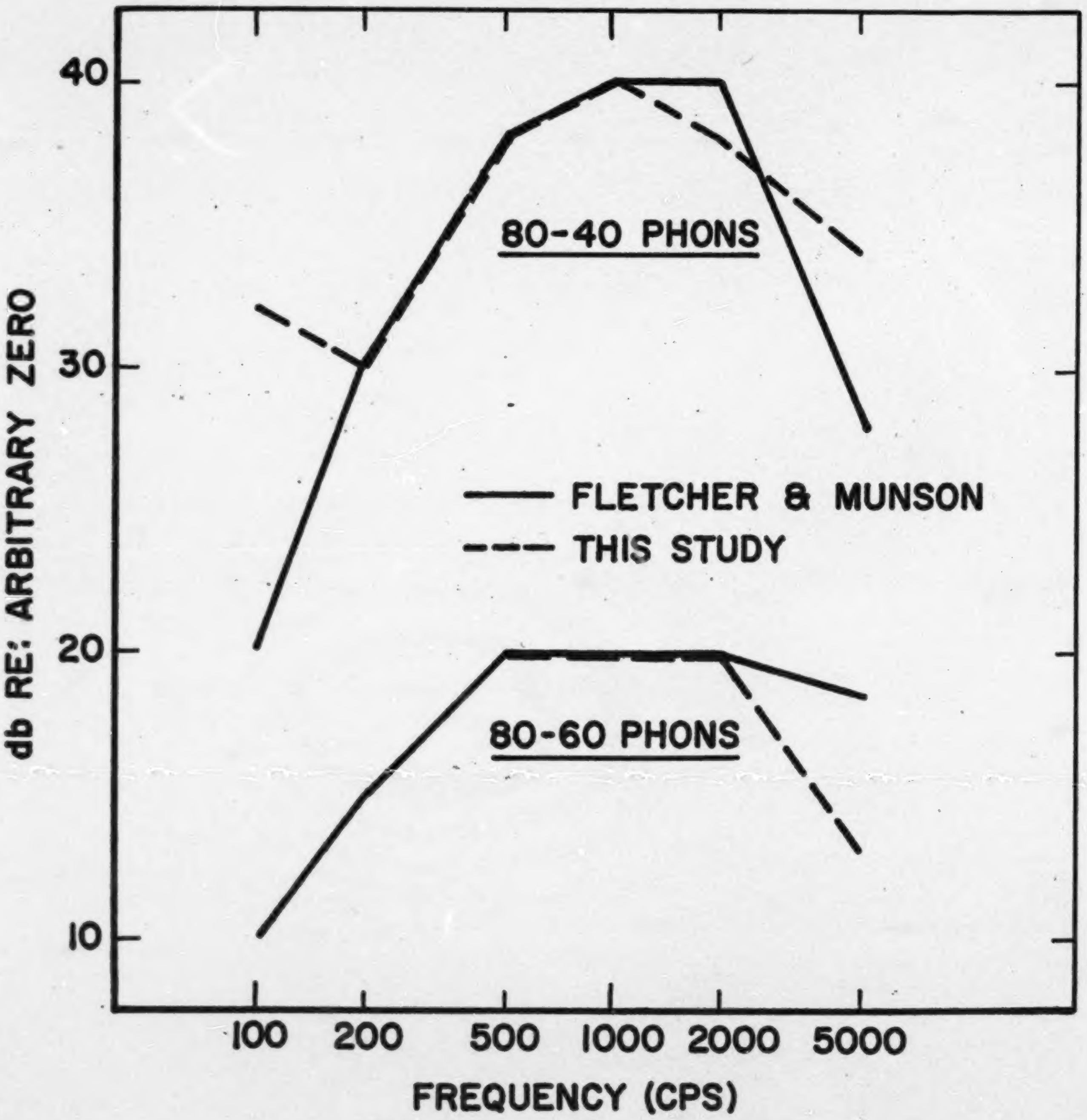

db RE: ARBITRARY ZERO
40
30
20
10
100
200
500
1000
2000
5000
FREQUENCY (CPS)
80-40 PHONS
80-60 PHONS
FLETCHER & MUNSON
THIS STUDY

Fig. 13

AN EFFECT OF NATIVE LANGUAGE AND MUSICAL TRAINING ON VOCAL PITCH MATCHING

H. L. Lane, J. R. Hunting, D. M. Brethower, and J. L. Kopp
The University of Michigan

An Effect of Native Language and Musical Training on Vocal Pitch Matching

H. L. Lane, J. R. Hunting, D. M. Brethower
and J. L. Kopp
University of Michigan

This study compares native speakers of English with and without musical training to native speakers of Thai, a tone language, in their accurracy and reliability when matching the fundamental frequency of voice to the frequency of a pure tone.

Method

Subjects

Fifteen males at the University of Michigan were employed: five were native speakers of Thai and had no musical training; five were native speakers of English without this training; five were native speakers of English who had received formal musical training (two were trained extensively in voice, the remaining three, with less tutelage, sang in public performances; one of these Ss was a composer).

Apparatus & Procedure

Five frequencies of a pure tone (100, 120, 140, 160, and 180 cps) were generated by an oscillator (Hewlett-Packard 200 AB), calibrated (Hewlett-Packard frequency counter 523 D), and recorded (Ampex 300) at a constant level, four times each in irregular order. Each of the 20 stimuli had a duration of 3 sec and was separated from the next by a 10-sec interval.

The S was seated in an audiometric room in front of a microphone

(Altec 633 A) connected to a tape recorder (Wolensack). The recorded stimuli were presented with a binaural headset (PDR-8) that had essentially flat frequency response over the pertinent range. The S was instructed to match the pitch of each stimulus, during the interval following it, by voicing the vowel phoneme /a/ with the same pitch and like duration.

The fundamental frequency of the recorded vocal responses was extracted by means of two band-pass filters (Krohn-Hite 310 AB) arranged in series (48 db/octave). The center frequency of the pass-band was adjusted for maximum rejection of harmonics by monitoring the filtered sinusoid on an oscilloscope. The center one-sec portion of the fundamental was sent to the frequency counter, which presented the number of cycles that occurred in one sec.

Results and Discussion

Figure 1 shows the relation between fundamental frequency of the vocal response and the frequency of the pure-tone stimulus, when Ss in the three groups were instructed to match the pitch of their voice to that of the stimulus tones. It does not appear that the speaker's native language or the degree of his musical training affected the slope of the pitch-matching function which was, for all three groups, nearly linear. When the pitch-matching function was drawn up for individual Ss within each group (not shown), it was found that the slope varied only slightly among Thai and "musical" American Ss, but varied widely among the non-musical Americans.

The variability in repeated matches to the same stimulus by each S, and the differences in matches among Ss, were both greatest for the American group without musical training. These two sources of variance have been combined and represented for all three groups in Fig. 1. Examination of the figure shows that the combined variance associated with each mean is greatest for the American group without musical training, less for the Thai Ss, and least for the group of Americans with musical training. These relations were verified by F-tests of the ratio of the variance in fundamental frequency for pairs of groups at each stimulus value.

Summary and Conclusions

Five Americans with musical training, five without, and five native speakers of a tone language, Thai, were asked to match the pitch of their voice, while producing a vowel, to that of a preceding pure tone. The pitch-matching functions for the three groups, relating mean fundamental frequency of the matching response to the frequency of the stimulus tone, were approximately linear and similar in slope. Americans without musical training had the greatest variance among the slopes of individual functions.

The variance for repeated matches by the same individual and for matches by different individuals showed major effects of native language and musical training. Both sources of variance in fundamental frequency were greatest for the Americans without musical training, less for the Thai Ss, and least for the Americans with musical training.

Figure Caption

Fig. 1. The relation between fundamental frequency of vocal responding and stimulus frequency when Ss with different language and musical backgrounds matched vowel pitch to stimulus pitch. Each point is the mean of 20 determinations, four by each of five Ss.

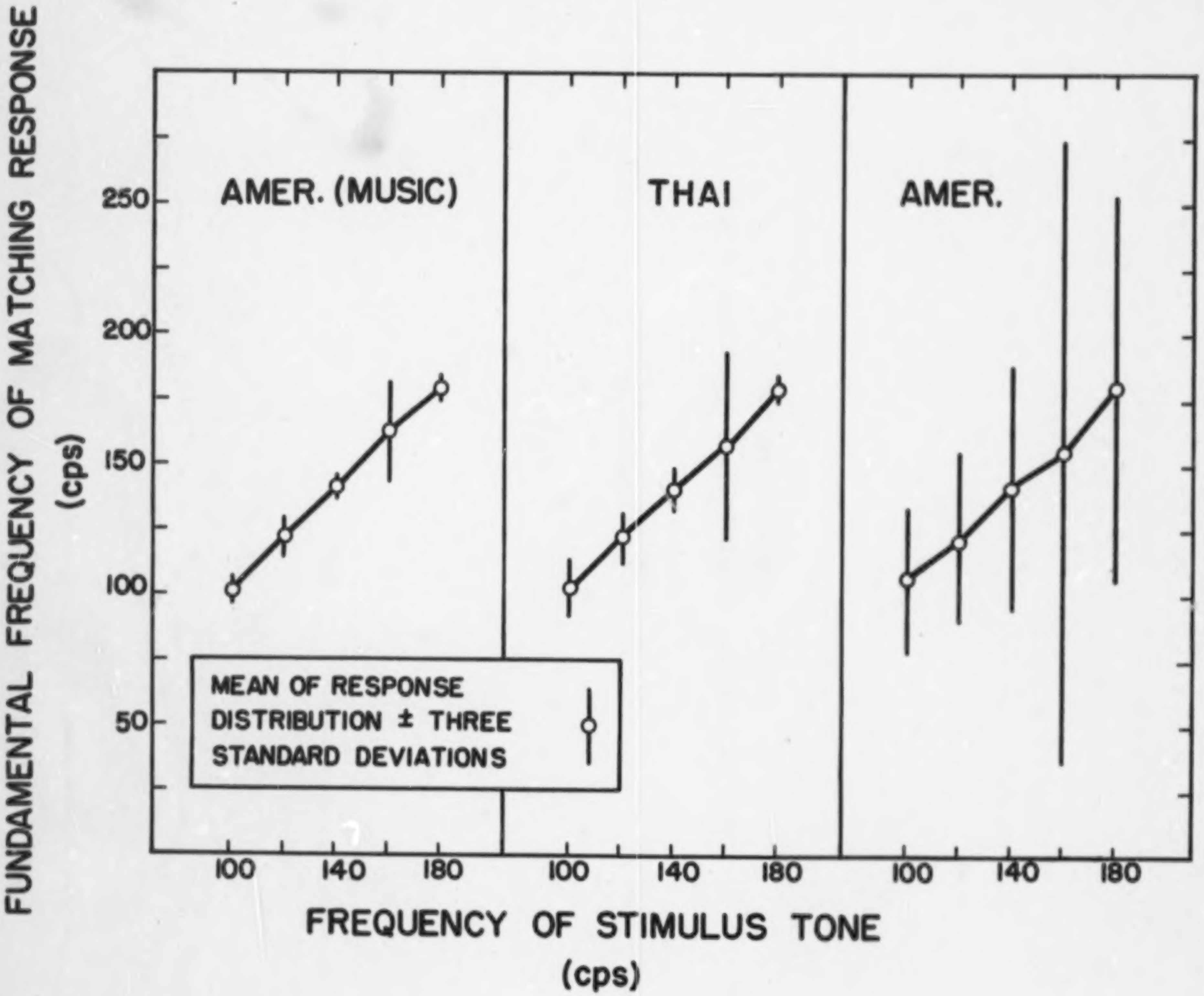

FUNDAMENTAL FREQUENCY OF MATCHING RESPONSE
(cps)
250
200
150
100
50
AMER. (MUSIC)
THAI
AMER.
MEAN OF RESPONSE DISTRIBUTION ± THREE STANDARD DEVIATIONS
100
140
180
100
140
180
100
140
180
FREQUENCY OF STIMULUS TONE
(cps)

Fig. 1

JUDGMENT OF PORTRAYED EMOTION FROM NORMAL ENGLISH, FILTERED ENGLISH, AND JAPANESE SPEECH

Ernest Kramer
The University of Michigan

JUDGMENT OF PORTRAYED EMOTION
FROM NORMAL ENGLISH, FILTERED ENGLISH, AND JAPANESE SPEECH

Ernest Kramer
The University of Michigan

TABLE OF CONTENTS

	Page
List of Tables	iii
I. Introduction	1
II. Method	13
III. Results and Discussion	20
IV. Summary	33
V. Appendices	35
A. The Complete Scripts Used by the American Actors for Portraying the Five Emotions	35
B. Japanese Test Passage	41
VI. References	42

LIST OF TABLES

Table — Page

1. Tape Rating Form — 17
2. Pretest Ratings of the Test Passage — 20
3. Per cent of Affect Ratings Assigned to Each of Five Portrayed Emotions — 22
4. Per cent of Affect Ratings Which Placed Each Portrayed Emotion in the Category Intended by the Actors — 23
5. Per cent of Affect Ratings Which Placed Each Portrayed Emotion in the Category Intended by the Japanese Actors — 29
6. Per cent of Judgments Which Agreed With the Portrayal Intended by the Good and Poor Judges Who Recorded the Test Passage — 30

I. INTRODUCTION

Review of Past Studies on:
The Judgment of Emotion from Non-verbal Properties of Speech

Most people hold informally the hypothesis that the non-verbal properties of voice carry information about a speaker and his feelings which his words alone do not convey. When we respond to someone with, "You don't really mean that!" or "Are you really sure?" we are often reacting to something more in his voice than the words--and, incidentally, we are providing some non-verbal, vocal cues of our own. There is an experimental literature in psychology which deals with the relationship between these non-verbal properties of voice and both stable personality traits and changing emotional states of the speaker (Kramer, 1962, a,b). Investigations of the relationship between voice and judgments of a speaker's changing emotional states must deal with the methodological problem of eliminating the effect of words. That is, how can the particular words spoken be eliminated or controlled so that the contribution of non verbal cues may be studied? The following review of studies bearing on the judgment of emotion from non-verbal properties of speech is organized according to how this problem was met. The studies are grouped together under the method which was used to control or eliminate the influence of verbal cues on the judgment of emotion.

Meaningless content

The chief forms of "meaningless content" have been numerals and letters of the alphabet; however, the earliest reported study of this type (Skinner, 1935) requested that the Ss merely say "ah." They were requested first to read a passage of emotional literature and listen to selected music which was designed to put them into a sad or happy emotional state. They then spoke, and the "ahs" of happiness showed higher pitch and greater force than those of sadness. Dusenbury and Knower (1939) asked a group of speech students and instructors to "try to feel the designated emotional state and to use a tonal code that would indicate their feelings" (p.67) while reciting the letters A through K. Each of 22 speakers tried to portray the 11 designated emotions. Eight sets of recitations were selected out of these 22 sets on the basis of pretests. Listeners were able to match all readings in these eight sets with the emotion which the speakers had tried to represent, with significantly greater than chance accuracy. In a later study (Knower, 1941) the speakers whispered the letters in an attempt to eliminate the effects of "tone"(the fundamental frequency of the voice is not present in a whisper). Recognition of the intended emotions was still better than chance. These studies unfortunately cannot indicate how closely these speakers, or any others, would approximate these same "tonal codes" when experiencing the emotions in real life situations. The fact that the "meaningless content" is produced without any verbal context surrounding it sets these studies still further away from these situations.

Similar material has been used for teaching psychotherapy (Thompson and Bradway, 1950). Two psychologists acted out a "therapeutic interview" in which they actually spoke only numbers, although with the inflections which a genuine exchange between patient and therapist might have. The two participants each listened separately to recordings of the interview and made statements about the "affective interchange." The two sets of statements were found to be significantly correlated with each other. Numerals also formed the only verbal content in a study by Pfaff (1954) in which an "experienced speaker" used numbers to express a variety of emotions. Among various groups of listeners, junior high school students of low socio-economic status did least well at identifying the emotions portrayed, while college oral interpretation students did best. A partial explanation of the results may be that the "experienced speaker" drew from the same stock of stereotypes and stage techniques with which the oral interpretation students were most familiar. The low ranking of the junior high school students of low socio-economic status suggests the hypothesis that the "tonal affect language" may be different, at least to some extent, for different classes in a society. Inexperienced readers from a Navy ROTC group were used by Black and Dreher (1955). They read a list of five syllable phrases in a manner aimed at simulating certainty or uncertainty. Listeners presumably of the same background as the speakers, were able to distinguish between the two types of reading with high reliability.

Recitations of the alphabet with varying expressions were used by Davitz and Davitz (1959 a, b) in work which focused on errors and relative ease in guessing various emotions. In their first study (1959a) all feelings were identified more consistently than chance alone would predict, but certain errors in identification showed significant consistency. When fear was mistakenly identified, it was most commonly taken for nervousness; love was most commonly misidentified as sadness, and pride for satisfaction. In their second study (Davitz and Davitz, 1959b) they selected from pretests two speakers who were particularly successful at communicating feelings through reciting the alphabet. These speakers each used the alphabet to express 50 different feelings. Thirty judges tried to match each recitation with the correct feeling. Ten judges rated each feeling from the list of 50, checking every other feeling on the list which was similar to it. A similarity score was based on the number of times a feeling was noted as being similar to any other feeling. Separate groups of judges rated the feelings on the list in order to derive "strength," "activity," and "valence" scores for them. Findings were: (1) accuracy of identification of feelings was correlated -.29 with similarity scores (significant at the .025 level of confidence), (2) the degree to which one feeling is mistaken for another is related to the subjective similarity of the two, (3) for pairs of similar feelings, the stronger tends to be communicated more accurately, and (4) no significant relationships appeared involving the activity or the valence scores. The authors note, however, that "since the relationships found were not high, the greater part of the variance in accuracy of communication is unaccounted for" (p.116).

The eight neutral sentences used by Pollack, Rubenstein, and Horowitz (1960a,b) may be regarded as "meaningless content" in the sense that they did not fit any context in which the emotional modes studied might occur. The sentences--e.g., "The lamp stood on the table"--were read in various modes under increasing signal-to-noise ratios. The authors found that recognition of the emotional words held up better under noise than did recognition of the particular sentences. Mode recognition was also possible with significant accuracy even the fundamental frequency was eliminated by having the sentences whispered. Effects of temporal sampling were also explored; some recognition of the modes was still possible with extremely short samples and with sections of the recorded samples removed at periodic intervals.

Constant content

Another technique for eliminating the effects on judgments of the words spoken has been to use the same set of words for all the emotions represented--words which are ambiguous as to emotional content, rather than numbers or separate letters. Like the use of meaningless content, this method requires actors--experienced or inexperienced--and thus may capitalize on the presence of stereotyped portrayals of emotions, which might not occur naturally. It has the advantage, however, of producing non-verbal cues in a verbal context: a more lifelike situation than the isolated letters and numbers in the studies noted above. Fairbanks (1940; Fairbanks and Hoaglin, 1941; Fairbanks and Pronovost, 1939) had experienced student actors read five passages, each marked by a different emotion. Listeners heard only a set of sentences that was common to all five passages; they were asked to identify the emotion being represented on the basis of these excerpts. Some of the actors seemed to provide much clearer vocal differentiation of emotion than did others. Measurable pitch differences (Fairbanks and Pronovost, 1939) and differences in duration of phrases (Fairbanks and Hoaglin, 1941) were found among the different emotions, using average measures from the various readings.

Ignoring content

Brody (1943), attempting to focus his attention away from the patients' words, has called attention to subtle variations in voice during the course of psychoanalysis. He presented several cased in which vocal changes seemed to mark transitions between major emotional stages in therapy. He regards vocal expression as a relatively safe way to act out hostile feelings during analysis. Verbal content may be "ignored" by measuring specific non-verbal properties of speech. Different aspects of speech rate and breathing rate were measured in a series of studies on noncontent properties of speech during psychotherapy (Goldman-Eisler, 1955, 1956a, 1956b). Rate of respiration and expulsion rate of syllables were correlated with a single psychiatrist's judgment of the emotional content of the patients' communications.

Experimental attention has been given to the role of silences and disturbances in speech as indicators of changing emotional states in psychotherapy (Mahl, 1956, 1959; Dibner, 1956). Mahl (1959) has sometimes referred to these as "expressive aspects of...speech," but they have chiefly been treated as disruptions in the speech process rather than as part of the simultaneous non-verbal accompaniment to spoken words which is the center of focus in the present review. Krause (1961) has shown that Mahl's measures are highly similar to those used by Dibner (1956) as indicators of anxiety. Krause and Pilisuk (1961), using measure from both Mahl and Dibner, found that "intrusive non-verbal sounds, mainly laughs and sighs" were the best speech indicators of transitory anxiety.

Linguistics has offered some schemes of analysis which consider non-verbal, or non-lexical, aspects of speech (Pittenger and Smith, 1957; Trager, 1958). These aspects, "vocalization and voice qualities together are being called paralanguage" (Tranger, 1958, p. 4). McQuown used these recently described categories in a detailed analysis of a single interview, where he characterized the affect involved; no independent measure of the affective aspects of the interview is reported. A highly detailed book-length analysis in these terms has been made of the first five minutes of a psychiatric interview (Pittenger, Hockett and Danehy, 1960). In a study by Eldred and Price (1958) four judges listened to tape recordings from various parts of intensive psychotherapy with a single patient and noted which linguistic and which emotional categories seemed to vary together. No cross-validation of the suggested relationships was reported. Some doubt has been cast on the utility of such approaches by a study using a larger number of judges (Dittman and Wynne, 1961). It was found that more traditional linguistic categories, such as juncture, stress, and pitch could be reliably coded but bore no relationship to the emotional state of the speakers. Paralinguistic categories were expected to be more emotionally relevant, but they could not be reliably coded.

Acoustical methods

Filtered speech. The studies noted here have eliminated verbal content by passing recorded speech through a low-pass filter designed to attenuate those higher frequencies of speech upon which word recognition depends. It is expected, in such studies, that many of the non-verbal properties of speech such as stress patterns and intonationpatterns based on fundamental frequency still remain. Little notice has been taken of the non-verbal information which is lost together with the verbal,although one acoustical study (Ochai and Fukamura, 1957) has shown that the upper as well as the lower overtones of speech contribute to the personal tone or timbre of a person's voice. Controls have also been lacking on the effectiveness of the filtering in eliminating content; the data on frequency and intelligibility which are most frequently quoted in the literature (Fletcher, 1953; French and Steinberg, 1947) do not deal with intelligibility in connected discourse.

Fifteen speech samples were rated for emotional content by a group which heard them on normal recording and separately by a group which heard them on recordings from which frequencies above 450 cps. had been removed (Soskin and Kauffman, 1961). The two sets of ratings showed significant agreement. Kauffman (1954) had a professional actor record two readings of a series of short speeches. In one reading, the actor read with an emotional expression which was appropriate to the words used in each speech, while in the other reading, he used an expression which was highly incongruent with the verbal content. The recordings were passed through a low-pass filter to remove the semantic content. One group of listeners judged the second series of speeches for incongruity by comparing the filtered recordings with typescripts of the speeches. Separate groups rated the typescripts alone, and the full range and filtered recordings. Kauffman classified the "meanings" of the rating categories into two main divisions: (1) expressive, "affect meanings relevant to the psychological state of the speaker," and (2) "manipulative...meanings relevant to the purposive behavior of the speaker." He found that both the vocal and verbal channels carry information about both the expressive and manipulative meanings in speech. There is, however, a tendency for the expressive function to be performed by the vocal channel and the manipulative by the verbal. Incongruence between vocal and verbal channels was reflected in greater heterogeneity of judgments, particularly in the judging of expressive meanings by those who heard only the filtered recordings. Heterogeneity of judgments was assumed to be a measure of ambiguity. There was, then, a consistent negative correlation between the degree of congruence of the vocal and verbal channels and the amount of ambiguity.

Starkweather (1955a, 1956a) sampled recordings of the 1954 Army-McCarthy hearings for three excerpts each of the voices of Senator McCarthy and Mr. Welch and filtered out content by attenuating higher frequencies. Twelve clinical psychologists showed high inter-judge agreement on rating emotional content, although they insisted that they had no confidence in their own ratings. The raters were then given a normal, unfiltered presentation of the excerpts and asked to rate them again. A comparison of the categories assigned to the filtered and unfiltered recordings indicates that Mr. Welch's voice was judged appropriate to the verbal content, while the Senator's voice was judged to be without variation.

Other acoustic techniques. Some possible experimental approaches to the role of duration and other acoustic properties of speech signals, which have not yet been directly used in voice and emotion studies, have been noted (Starkweather, 1961). These may offer better ways of quantifying some of the dimensions of speech which change with changing emotions. Hargreaves and Starkweather (1961a) present a case study where judges were able to use certain parts of speech spectrograph records to identify changes in a patient's vocal behavior which had been considered significant by her therapist. The validity of the "machine method" of identifying emotionally significant vocal changes still rests on the validity of the skilled listener who sets the criterion dimensions for it, but the authors feel that the method offers a great saving in effort over having a skilled listener consider separately every section of vocal behavior in an interview. Using predetermined features of the spectrogram also avoids the effects of fatigue and the learning of wrong cues which might mar the judgments of the skilled listener alone. The authors point out that much information is present in the spectrogram, and a different selection of dimensions might have served as well as theirs for finding correlates to emotional changes in the patient. It may be necessary to use different dimensions for different individuals, as Krause (1961) has found different behavioral measures of vocal behavior to be important for different subjects in identifying anxiety.

The Present Study

The present study investigated problems in four areas of the general field of judging emotion from non-verbal properties of speech. An attempt has been made to fill certain gaps in the experimental literature reviewed above.

Relationship between verbal and non-verbal cues in judging emotion

Only one previous study has considered the effect a subject's emotional judgment of the words of a passage might have on his judgments when hearing the passage spoken with various emotional expressions (Kauffman, 1954), and that study used extremes of incongruity between verbal and non-verbal cues. The present study predicted that even with relatively ambiguous material, a subject's emotional rating of the words alone would be predictive of his ratings of non-verbal cues when presented in association with the passage. This prediction stems partly from the expectation that a subject has certain needs and emotions of his own which remain relatively constant and which will color his reaction to a passage under varying conditions. This seems consistent with the extensive literature on clinical projection and need-influenced perception (Atkinson, 1958).

Hypothesis 1. An individual's judgment of the emotional content of an ambiguous verbal passage will tend to make him give the same judgment of its content when he hears it spoken with the various, relatively unambiguous, emotional expressions.

Affective information in filtered speech

As noted in the preceding review, past studies using speech with the high frequencies attenuated have generally not considered how much affective as well as semantic information was lost in the filtering process. Since individuals may vary in their relative ability to judge emotion in speech under normal and filtered conditions, investigating the possible loss of affective information requires that each subject hear the same sample under the varying conditions; that is, that each subject serve as his own control. In past studies however, when the same subjects have heard both filtered and unfiltered presentation of the same material, it seems to have assumed that differences between the conditions were due only to the influence of the semantic information provided by the words (Starkweather, 1955a, 1956a). The present study, which controlled the semantic information (See Chapter II, METHOD), predicted:

Hypothesis 2. Removing those higher frequencies of speech upon which word recognition depends will significantly reduce the non-verbal, vocal cues for emotion.

Non-verbal, vocal cues across languages

No previous study appears to have investigated the judgment of emotions from speech in a language with which the listener is unfamiliar. Japanese was selected as a foreign language with which to investigate this problem in the present study. It is a language whose vocabulary is virtually unknown to the university student community in the United States, the population from which the present subjects were drawn.

Japanese was studied in a recent investigation of cross-linguistic phonetic symbolism (Miron, 1961). Nonsense syllables, composed of phonemes which occur in both Japanese and American English, were presented on tape recordings to Japanese and American subjects. The two groups both rated the syllables on the semantic differential (Osgood, Suci and Tannenbaum, 1957), and marked common tendencies in their ratings were interpreted as evidence for the existence of cross-linguistic phonetic symbolism. Since the vocal expression of the phonetician who recorded the syllables may have been influenced by his own reaction to the syllables, it seems possible that non-verbal vocal cues may have been produced which were similarly interpreted in both American and Japanese.

Hypothesis 3. American listeners will be able to categorize the emotions being portrayed in Japanese by Japanese speakers with an accuracy significantly better than chance.

Relationship between perception and transmission of affect in speech

No experimental work seems to have been reported on individual differences in the ability to comprehend and convey emotion through voice. In verbal language one expects some relationship, albeit an imperfect one, between the ability to understand and the ability to speak a given language. If the non-verbal cues constitute a consistent "language" of their own, it seems possible that those who understand the "language" better than others can also speak it better.

Hypothesis 4. Individuals who are most successful in judging emotion from the speech of others will be able to convey emotion more recognizably in their own voices than will individuals who are relatively poor at judging emotion from speech.

II. METHOD

The test passage

A common set of words was chosen to represent each of the emotions which were considered in the study. This technique of holding content constant was used by Fairbanks (1940; Fairbanks and Hoaglin, 1941; Fairbanks and Pronovost, 1939) to remove the effect different sets of words would have in providing cues about different emotions. The passage, modified from that used by Fairbanks, reads:

> There is no other answer. You've asked me that before.
> My reply has always been the same, and it will always
> be the same.

As noted in the preceding review of the literature, eliminating effects of content by holding it constant requires the use of actors. This study then, like previous ones, is concerned with the judgment of portrayed or simulated emotion.

The test passage was first presented in typescript to 72 students in an advanced undergraduate psychology course. They were given a list of emotions, similar to that which would be given to Ss in the principal part of the study, and asked to check which emotion best described the passage. Half of them had the emotions listed in alphabetical order on their check sheet, while half had the order reversed. From the results of this pretest rating (See next chapter), the final form of the check sheet was determined.

Recordings by American actors

The test passage was embedded in five different scripts, each aimed at portraying a situation which would give rise to speech expressing one of the emotions: anger, contempt, grief, indifference, love.[1] The scripts were recorded by six male students from the University speech department and by the experimenter, all of whom had experience in radio acting. The actors were told what emotion was to be protrayed with each script. If the actor or the experimenter or a radio director who participated were dissatisfied with any reading and felt it could be improved, then another recording of that particular script was made by the actor.

---

[1] The complete scripts are given in the appendix.

Editing and selection of tape recordings

The test passage alone, excerpted from 43 of the actor recordings, was presented in random order to 25 students in an introductory psychology course. They were asked to check on their rating sheets which one of the five emotions listed--anger, contempt, grief, indifference, love--best fitted each reading of the passage. Twenty-two readings were selected for the remainder of the study on the basis of these ratings. In each of the 22, the emotion most frequently checked by the raters was the same as that the actor had been trying to portray. There were five readings each of anger, grief, and indifference, three readings of contempt, and four of love.

The filtered speech tape

The 22 selected passages were copied onto another tape through a low-pass filter. Two Ampex tape recorders (Model 350) were used, and two Kronhite band-pass filters (Model 310 A-B) were connected in series between them. Both filters were set to pass frequencies below 400 cps., and together they gave a filter attenuation of 48 db per octave. The filtered versions of the passage were arranged on the tape in random order. Preceding them was the following introduction:

> You will hear shortly some tape recordings which have been passed through a special electronic filter. This filter removes certain parts of the sound so that it becomes virtually impossible to recognize the words that are being spoken on the recording. Much of the sound, however, does

> remain. You will be asked to judge from these filtered recordings what emotion was being expressed in the original recording.
>
> Since the filtered sound may be a rather strange one at first, you will hear now a few samples of such filtered recordings. These are not among those you will be asked to rate, but are merely to give you an idea of the type of things you will hear. Each of the following filtered recordings is followed by an unfiltered presentation of the same material.

Three brief excerpts from role-play situations unrelated to the present study were then played in filtered and unfiltered recordings. Next, the tape alerted the listeners that the recordings to be judged were next, and that these would be presented in filtered recordings only, and instructions were given for the use of the rating sheet. A number was announced before each reading of the passage, so the Ss could identify the passages on their rating sheets. The first four readings were repeated at the end of the tape. This initial preset was not used in testing the hypotheses. It gave Ss a further chance to become accustomed to the judging task, and it also provided a means of estimating the reliability of ratings.

Recordings by Japanese actors

The five scripts used by the American actors were translated into Japanese by three Japanese graduate students.[2] Two of the translators were linguistics students in Japan, and one had taught linguistics. The translators recorded the test passage, giving three readings each for the five emotions they were trying to portray. They were asked to speak the lines as they would in communicating these emotions to a Japanese listener who knew no English. They then listened to their own recordings and judged together which of the attempts at portraying each emotion came closest to meeting the criterion of a convincing portrayal of that emotion as it would be expressed naturally in Japanese. Recordings were rated from 1, most successful, to 3, least successful.

---

[2] The Japanese translation of the test passage is given in the appendix in English transliteration and phonetic alphabet.

The unfiltered tape

Twenty-two of the Japanese recordings of the test passage were selected, distributed among the five emotions as were the 22 English excerpts. Those Japanese recordings given a rating of "1" were selected first, and the remainder taken randomly from among those given a rating of "2" by the Japanese speakers. A tape recording was prepared in which the English recordings alternated with the Japanese. The alternation of E.-J.-E.-J.-etc. was designed to avoid practice effects on the general task of listening and judging from affecting one language more than the other, so that differing success on the two languages could not be attributed to such practice effects. As with the filtered tape, the final four passages were presented first in a "pre-set" at the beginning of the tape. A number was announced before each reading from 1-48, including the pre-set.

The tape rating booklets

Rating forms were prepared as shown in Table 1, "Tape Rating Form." These were stapled into booklets of 48 each, for the English-Japanese tape, and 26 each for the filtered English tape; the pages were not numbered. Since each rating form constituted a separate page, Ss could not easily keep track of the ratings given to past excerpts as they rated new ones.

The typescript booklet

A booklet was made up of the test passage and 11 other ambiguous phrases. Each phrase was on a separate page, together with a rating form like that which the Ss would use to rate the recorded passages (See Table 1). The additional passages were added to prevent Ss from easily and immediately recognizing the test passage when they heard it, following the typescript ratings, on filtered recordings. The test passage was presented before the additional ones, so that the rating given it would be uninfluenced by those given the other passages. A sheet asking for family, school, and medical data about the *S* was also included in the typescript booklet, before the ambiguous phrases.

Table 1

Tape Rating Form

Number______________

__________Love

__________Indifference

__________Grief

__________Contempt

__________Anger

Barely noticeable emotion ____ Extreme emotion

Procedure during the rating sessions

The tape recordings and typescripts described above were rated by each of 27 male students, who were paid for participating in the one hour session.

1. *Ss* were given the typescript booklet. They filled out the data sheet and rated the 12 phrases.

2. The audibility of pure tones of frequency 4000, 3000, 2000, 1000, 500, 250, and 125 cps. at 20 db (SPL) was determined for each ear of each of the *Ss*. The criterion for rejection of an *S* was two failures to respond to tones anywhere in the series. No *Ss* had to be rejected.

3. The filtered recordings were presented with a Magnecord (Model PT6-BA) over four binaural headsets, containing matched earphones (Grason-Stadler, Model TDH 39), which had a flat frequency response (± db) from 100 to 3000 cps. The headsets were wired in a series-parallel network so that the sound

levels in each were equal (±3 db), and their combined impedances matched the source impedance of the Magnecord. The sound pressure level in the phones centered around 72 db, with a range of approximately 20 db. The tape recording contained instructions for the rating procedure:

> A number will be announced before each one. Write that number in the space provided. After hearing each recording, check the one emotion on your rating sheet which most nearly describes the emotion present. Then, check on the scale the amount of emotion you feel is present. There will be about five seconds between each recording and the number announcing the next one.

The passages were rated in the tape rating booklets described above. Ss were not told how many readings to expect; at the end of the series an announcement was made on the tape that this part of the experiment was over and the S might remove his earphones.

4. About five minutes were allowed between presentation of the filtered tape and the tape recordings which followed. The Ss were asked, if they had not mentioned it spontaneously, whether they had been able to recognize the words spoken anywhere on the filtered tape.

5. The unfiltered Japanese-English tape was presented and rated under conditions identical to those used for the filtered speech recordings.

Recordings by the listener-actors

Following the sessions described above, the two Ss who had been most successful in placing the unfiltered English readings in the category intended by the actor, and the two who had been least successful recorded the same five scripts which the actors had used. They rehearsed by themselves for 15 minutes, recorded the scripts, rested for a few minutes, and then recorded them again. No direction designed to influence their readings was given.

The recordings of the test passage in the second series of recordings made by each of the four listener-actors were copied in random order onto another tape. This tape was then played for 19 of the same subjects who had participated in the original listening sessions. The listening and rating conditions were the same as that used for the previous unfiltered tape.

III. RESULTS AND DISCUSSION

Pretest of the test passage

The ratings of affect assigned to the test passage by 72 students who saw it only in typescript are presented in Table 2.

Table 2

Pretest Ratings of the Test Passage

		Order	
	Alphabetical	Reverse	Total
Anger	12	14	26
Contempt	11	9	20
E, F, G*	3	3	6
Indifference	7	3	10
Love	3	7	10

*Embarrassment, fear, and grief were all infrequently chosen. Their frequencies were combined in order to permit a chi square test for order effects.

Four emotions were chosen fairly frequently: anger, contempt, indifference, and love. These four emotions together with one of the infrequently chosen ones, grief, were selected for the multiple choice rating form. (Table 1)

To determine if the order of listing the emotions on the rating sheet affected the emotions chosen, the number of people who chose each emotion under the two conditions was compared with the null hypothesis that there would be no significant difference in the number of people choosing each emotion under the two conditions. A Chi square test yielded a chi square ($p > .30$) which did not require rejection of the null hypothesis.

Judgment of emotion from unfiltered, English speech

The results of judging affect from the unfiltered, English recordings is given in the top matrix of Table 3. All the portrayed emotions were most frequently judged as the actor had intended them. The most frequent confusion between two emotions, that between grief and love, is the same as earlier studies had found (Davitz & Davitz, 1959a; Fairbanks & Pronovost, 1939). Unlike this error, the relationship between anger and contempt is not symmetric; that is, anger may be judged as contempt, but contempt is only rarely judged as anger. Some emotions seem harder to judge than others. Readings portraying grief and love were less frequently placed in the intended category, partly because of the confusion between these two, than were the portrayals of other emotions. It would be unsound, however, to attempt to generalize such differences among emotions very far. No statement about the "purity" or "representativeness" of the emotions in the recorded samples can be made.

Not all actors were equally successful in portraying emotion--a finding also reported by Fiarbanks and Pronovost (1939). The percentage of judgments which placed each actor's portrayals in the intended category is given by the underlined numbers in Table 4. The total per cent judged correctly--that is, placed in the intended category--is only roughly comparable between actors, because not all actors were represented in all emotions. Individuals apparently vary in their ability to portray different emotions. An actor may be relatively successful with one emotion and comparatively unsuccessful with another. If such differences also occur in spontaneously expressed emotion, it would help to explain the problem presented by persons who seem unable to make certain feelings clear to others although they are not generally stolid or withdrawn.

Hypothesis 1 predicted that an individual's judgment of the emotional content of an ambiguous verbal passage would tend to make him give the same judgment of its emotional content when he heard it spoken with various,

Table 3

Per cent of Affect Ratings Assigned to Each of Five Portrayed Emotions

Emotion Intended by the Actor

Emotions As Judged by the Listeners

a. ENGLISH (unfiltered)

	Anger	Contempt	Grief	Indiff.	Love
Anger	74	3	2	1	2
Contempt	26	85	5	16	14
Grief	0	3	58	4	17
Indiff.	0	9	4	76	11
Love	0	0	31	3	56

b. FILTERED ENGLISH

	Anger	Contempt	Grief	Indiff.	Love
Anger	77	6	0	6	0
Contempt	23	48	3	11	12
Grief	0	4	71	7	20
Indiff.	0	36	5	63	19
Love	0	6	21	13	48

c. JAPANESE

	Anger	Contempt	Grief	Indiff.	Love
Anger	67	0	0	2	1
Contempt	29	20	0	21	7
Grief	0	17	90	2	3
Indiff.	3	52	1	73	52
Love	1	11	10	2	38

Table 4

Per cent of Affect Ratings Which Placed Each Portrayed Emotion in the Category Intended by the Actors*

	Intended Emotion					
	Anger	Contempt	Grief	Indiff.	Love	Total
DC	56 100	--	63 59	--	--	59 80
FK	96 93	--	--	--	74 22	85 57
HB	44 44	--	67 74	93 78	--	68 65
EX	93 85	85 52	93 96	85 59	59 59	84 71
IS	81 59	--	15 48	85 81	--	60 63
JR	--	89 44	--	67 52	--	78 48
BS	--	85 48	52 74	48 44	46 56	56 56

*Underlined numbers are the per cents for unfiltered, English recordings; the other numbers are for the filtered recordings.

relatively unambiguous, emotional expressions. The emotion which each S assigned to the typescript of the test passage was compared to the emotion which he used most frequently to match the unfiltered, English recordings of the passage. The modal category was the same as that assigned to the typescript for eight of the 27 Ss. A chi square test was used to compare this result to that expected if the agreement were purely on the basis of chance: an expected agreement of one out of five. The obtained chi square had a probability of between .30 and .20. Thus, although the results were in the direction predicted by the hypothesis, they were not statistically significant.

The criterion for agreement between the typescript rating and the modal category in the listening judgments was a stringent one. When an S had a multi-modal distribution of judgments on the recordings--if, for example, he used both love and grief with high and equal frequency--the mode was considered as not matching his rating of the typescript, even though one of the modes might be the same. A stronger influence might have been exerted by a less ambiguous and more emotionally loaded test passage, but this cannot be judged from the present study.

Two of the first four readings on the unfiltered tape--the "pre-set" readings which were repeated later in the tape--were in English. The emotions assigned to these were compared with the emotions assigned by the same Ss to the recurrence of these readings at the end of the tape. Seventy-eight per cent of the judgments were the same for the two presentations of the same readings. Only 20 percent agreement would have expected on the basis of chance alone; a chi square test shows that the obtained result differs from the chance expectation at well beyond the .001 level of confidence.

Judgment of emotion from filtered speech

Intelligibility of filtered speech. Studies using low-pass filtered speech to remove semantic content have varied both in the characteristics of the filters used and the amount of information supplied about them. Soskin and Kauffman (1961), for example, report using bands of 150-450 cps. and 150-550 cps. with a filter attenuation of approximately 65 db per octave, while Starkweather (1956a, 1961) reports that "frequencies above 300~were severely attenuated" by a "cruder electronic filter" than that used by Kauffman (1954). No measure of intelligibility of the remaining speech sounds is reported, although lack of any recognizable connected discourse is assumed. The present study used a filter which attenuated frequencies above 400 cps. with 48 db attenuation per octave, as an approximation to previous filtering techniques. No measure of intelligibility was included in the design, but several of the Ss spontaneously reported that they had been able to recognize the words on the filtered tape. Closer questioning of these and other *Ss* revealed that none had been able to recognize the words on the *first* few readings of the filtered test passage, but nearly all of them were able to do so by the time all 26 readings had been played. There were differences among *Ss* as to when in the series they were able to understand the words. Most reported that it seemed as if the degree of filtering was decreasing; two *Ss* asked *E* if the tape had been arranged so the readings would get progressively clearer. Apparently, as the information contained in filtered speech is presented again and again, more and more of it is utilized. There were no more cues in the 26th excerpt than in the fourth, but by this time, *Ss* were able to use more of the same cues which each excerpt provided. Because all excerpts contained the same words, no conclusion can be drawn from the present study relating intelligibility of filtered speech to increasing familiarity of the speech sample.

Hypothesis 2 predicted that removing those higher frequencies of speech upon which word recognition depends would significantly reduce the nonverbal, vocal cues for emotion. It was expected that the *Ss* would be less successful with the filtered than with the unfiltered speech in assigning to the readings of the test passage the emotion intended by the actors. The differences between success on filtered and unfiltered speech were found for each *S*. These 27 differences were arranged in rank order without regard to sign. The original sign was then restored to the rank, and the Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) was used to determine whether the result was significantly different from a null hypothesis that the number of *Ss* who were more successful in judging filtered speech was not significantly different from the number who were less successful. The obtained result did not permit rejection of the null hypothesis, therefore Hypothesis 2 of this study was not supported.

Previous studies using filtered speech (Cohen, 1961; Kauffman, 1954; Soskin and Kauffman, 1961; Starkweather, 1955a,b, 1956a,b, 1961) have assumed that filtering had no effects on judgments beyond the removal of semantic information. Data from the present study indicate that this was an erroneous assumption. Although there was no significant difference between the number of Ss who were more successful in judging affect from filtered speech over normal speech and the number of Ss who were less successful, the success of any particular S with one condition of speech told nothing about what might be expected from him under the other condition. The correlation between scores on the filtered and unfiltered presentations of the readings was only +.16 (Spearman rho, $p > .40$). The patterns of errors and successes for different emotions also shows some differences in filtered speech, as may be seen by comparing the top and middle matrix in Table 3. Grief and love are still frequently confused, and anger is mistaken for contempt; but portrayals of contempt are frequently rated as indifference. The recordings portraying grief are somewhat more successfully recognized from filtered than from unfiltered recordings, though the difference is not a large one. Reference to Table 4. shows that the relative success with which different actors conveyed different emotions was not the same for filtered and unfiltered speech--still further evidence that judging filtered speech is a different sort of task than judging emotion from normal recordings. Presumably, part of the differences between actors in conveying emotion under the two conditions is due to the fact that information about emotion may be carried by different vocal properties in different speakers. Those actors whose portrayals could be recognized as frequently, or even more frequently, under the filtered condition may be those who convey emotion chiefly through properties which remain in filtered speech; e.g., changes in fundamental frequency, changes in amplitude of the fundamental, changes in tempo.

Some clue to the difference between judging filtered and normal recordings may be offered by the fact that success in judgments of Japanese speech correlated higher with judgments of filtered speech (+.34, $p < .05$), than did success with normal, English speech (+.16, $p > .40$). The common element between Japanese and filtered speech is probably not semantic unintelligibility, since, as noted above, most of the filtered readings were intelligible to many of the *S*s. Perhaps it is that they both demand judgments under "strange" or "unusual" conditions. There may be individual differences among listeners as to how effectively they can judge stimuli for emotion under such conditions.

The emotions assigned to the four pre-set readings were compared with the emotions assigned by the same *S*s to the recurrence of these readings at the end of the tape. Forty-six per cent of the judgments were the same for the two presentations of the same readings. This result, like that for unfiltered English, differs significantly from chance agreement ($p < .001$). It should be noted that pre-set--repeat reliabilities cannot be compared for size between the filtered and unfiltered readings, as different readings of the test passage were used in the pre-sets in the two conditions.

Judgment of emotion from Japanese speech

Hypothesis three predicted that American listeners would be able to categorize the emotions portrayed in Japanese by Japanese speakers with an accuracy significantly better than chance. Since the present study required *S*s to select the one matching emotion from five choices, one out of five of the matchings would be correct on the basis of chance along. A comparison, by chi square, of the obtained number of correct matchings with the

chance expectation shows that the matching was significantly better than chance ($p < .001$): Hypothesis 3 is supported. The over-all pattern of successes and errors in judging may be seen in the bottom matrix in Table 3. Grief proved easiest to judge, with indifference and anger next. Love was mistaken for indifference more times than it was correctly judged, although the number of correct judgments was still greater than chance. Contempt was least often judged correctly; it was the one category where correct matchings were not significantly greater than incorrect ones.

The Japanese actors reported that grief was the easiest emotion for Japanese to express openly, anger and contempt next easiest, and love most difficult; "indifference" they regarded as fairly easy to portray, but did not class it as an "emotion" with the others. This report fits the results for grief and anger well, but leaves unexplained the lack of success in judging contempt. One of the Japanese actors was asked to describe the manner in which he expressed contempt on the reading which seemed most successful to him. He explained that he did so, as Japanese often did, merely by raising the pitch of this voice. Even if American listeners might recognize the emotional significance of this raised pitch in context (which is by no means certain), it is impossible clearly to recognize it from the recordings used in this study where there was no baseline of natural voice pitch to judge the contempt excerpts against.

The errors in judging love were chiefly due to the portrayals by one actor, "Tan," as may be seen from Table 5. His portrayals of love were judged as such by American listeners only seven per cent of the time, while the portrayals by "Tag" were judged successfully 69 per cent of the time. No such marked difference between the portrayals was noted, however, by the actors themselves. Taken together with the results for contempt, this suggests that not all non-verbal, vocal expressions of emotion which are recognizable between native Japanese can be recognized by American listeners. Where there are a number of ways of expressing an emotion, such as love, in Japanese, only some of these may be recognizable by Americans. Some emotions, such as contempt, may be expressed in Japanese chiefly by means which are unrecognizable to American listeners.

Table 5

Per cent of Affect Ratings Which Placed Each Portrayed Emotion in the Category Intended by the Japanese Actors

Actor			Intended	Emotion		
	Anger	Contempt	Grief	Indiff.	Love	Total
Tan	69	37	74	65	7	49
Tag	83	--	89	85	69	81
Ima	33	11	98	63	--	52

The correlation between success in judgments on the Japanese readings and success in judgments on the unfiltered, English readings was +.29 (Spearman rho, $p > .05$). When the correlation was computed without consideration of the attempts to judge Japanese contempt, however, it rose to +.41 ($p < .05$). Apparently, it is possible to make some predictions from an American listener's success in judging emotion from English speech as to how he will do in judging Japanese speech; the accuracy of the prediction depends partly on which emotions are involved.

Two of the first four readings on the unfiltered tape--the pre-set readings-- were in Japanese. The emotions assigned to these were compared with the emotions assigned by the same Ss to the recurrence of these readings at the end of the tape. Fifty-nine per cent of the judgments were the same for the two presentations of the same readings. This result, like those obtained for both filtered and unfiltered English readings, differs significantly from chance agreement ($p < .001$).

Differences Among Raters in the Successful Judgment of Emotion

Hypothesis 4 predicted that individuals who are most successful in judging emotion from the speech of others will be those who are also able to convey emotion more recognizably in their own voices than those individuals who are relatively poor at judging emotion from the speech of others. Table 6 shows that one of the two poorest judges of emotion (DGK) was apparently able to convey emotion better than one of the two best judges (RNG). The hypothesis, then, is not supported. It was noted, however, that when an error was made in matching the portrayals by the least successful judges, the error was usually to call the portrayal "indifference." Errors in judging the portrayals of the better judges were,

Table 6

Per cent of Judgments Which Agreed With the Portrayal Intended
by the Good and Poor Judges Who Recorded the Test Passage

	Good Judges		Poor Judges	
	HSN	RNG	DGK	RVB
Anger	67	11	59	19
Contempt	52	48	4	4
Grief	33	4	44	7
Indiff.	11	41	41	56
Love	22	19	4	0
Total %	37	24	30	17
	31		24	

on the other hand, more evenly distributed. The total number of times indifference was used to match a reading was compared for the portrayals by the two most successful and the two least successful judges. A chi square test was used to compare the obtained results with the chance expectation that both sets of readings would have equal numbers of indifference ratings. Indifference was significantly assigned more often to readings by the least successful judges ($p < .001$). It would seem that when the better judges try to portray an emotion, they do manage to convey some feeling of affect or involvement even if they cannot portray one emotion as clearly different from another; while when the poorer judges of emotion cannot sharply convey the particular emotion called for in their portrayals they tend to sound merely indifferent. The findings on these four speakers must, of course, be somewhat tentative due to the small N.

It was noticed that the two most successful judges had no siblings, while the least successful had one or more siblings. The difference between success in judgments by all Ss with and without siblings was compared by the Fisher exact probability test (Siegel, 1956), with the expectation that there would be no difference between the two groups of Ss. Ss without siblings did better at judging emotion from unfiltered English speech at the .05 level of confidence or better. When total success in judgment was taken for the filtered and unfiltered English recordings together, the greater success of Ss with no siblings was significant at, or better than, the .01 level of confidence. There was no significant difference, however, on the filtered recordings alone or on the Japanese recordings.

A tentative after-the-fact explanation for this unexpected finding is that only children may have to depend more on communication from adults than do children with siblings; in the years before a working vocabulary has been learned, it may be the non-verbal vocal cues upon which social behavior is based. The fact that this greater success of only children in judging emotion does not extend to either Japanese or filtered English adds further support to the suggestion that common factors occur in judging Japanese and filtered English recordings which do not occur in judging normal, English speech (See the discussion above under "Judgment of emotion from filtered speech"). This presumed ability to make judgments of emotion under unusual conditions would, then, be unrelated to whether Ss had siblings or not.

Future research

The present study dealt with four neglected areas in the study of judgments of emotion from non-verbal properties of speech: the effect of verbal cues on judgment of non-verbal cues, the differences between judging filtered and unfiltered speech, judging non-verbal cues from a foreign language, and certain sources of individual differences in success in judging emotion from speech. Two possible extensions of this work seem clear. It is hoped that the Japanese and English recordings can be presented for judgment to listeners who know only Japanese and to bilinguals from both Japanese and American cultures. The recording and judging conditions used in this study dealt with portrayed emotion; it is hoped that further research can be done on naturally occuring emotional speech.

IV. SUMMARY

A critical review is presented of experiments in psychology which deal with the judgment of emotion from non-verbal properties of voice. Various methods have been used to eliminate the effect of the particular words spoken on judgments. Among these methods have been passing the speech samples through electronic filters which attenuate those higher frequencies of sound upon which word recognition depends, and using the same words to express different emotions.

The present study used the method of keeping words constant for different emotions. Recordings were made by actors who portrayed five different emotions from five scripts, each of which contained a test passage of exactly the same words. The test passage was also recorded by Japanese actors, in Japanese, who tried to portray the same five emotions: anger, contempt, grief, indifference, and love. Listeners heard and made judgments of emotional content from normal recordings of the test passage in English, recordings of the English test passages where high frequencies had been attenuated, and recordings of the test passage in Japanese.

Listeners are able to match normal recordings of portrayed emotion with the emotion intended by the actor significantly better than chance. The emotion which each S used most frequently for matching the readings showed a slight, but not statistically significant, tendency to be the same emotion which he had used to match a typescript of the ambiguous test passage.

The over-all percentage of correct matches was approximately the same for normal and filtered recordings of the portrayals by American actors, but there was no significant correlation for rater judgments between the two conditions. The previous literature has assumed that the only difference in judgment between normal and filtered speech was due to the absence of words in the latter. The present study shows that this is not true; knowledge of how a person judges normal speech permits no prediction of how he will judge filtered speech. Ss judgments on filtered speech did show a small but statistically significant correlation with their judgment of the portrayed emotion in Japanese speech. This suggests a common element of ability to judge emotion under unusual or abnormal conditions.

American Ss were able to match emotions correctly with those portrayed in Japanese by Japanese actors. Not all emotions proved equally easy for American listeners to recognize in Japanese. Furthermore, some emotions may be portrayed in Japanese in more than one way, not all of which are recognizable to Americans. No cross-linguistic study of judging emotion from non-verbal properties of speech appears to have been reported in the previous literature.

Two Ss who were particularly successful in matching unfiltered English portrayals with the intended emotion and two who were particularly unsuccessful recorded portrayals of the same emotions. Listeners were not better able to judge the portrayals of the successful judges. Ss without siblings were more successful in judging emotion from portrayals by American actors than were those with siblings. This relationship was not true for judgments of Japanese or filtered English speech.

APPENDIX A

The Complete Scripts Used by the American Actors for Portraying the Five Emotions

ANGER

The candidate for senator has been asked again by a group of powerful members to compromise his stand on a key issue, in order to win more votes.

THE CANDIDATE: Look! You're wasting your time! You've come here with that compromise offer three times, and three times I've said, "No." Well, it's still "No!" Yes, I realize you're among the men who helped me get this far in politics, but it was a hell of a lot of honest voters who helped me get here, too. I will not be intimidated by you. I've had enough of your honeyed phrases and enough of your subtle threats. Now you can go! And one more thing before you go--wipe off that "we understand each other" smile! We don't understand each other. I don't want to have to understand your brand of politics or you. Damn the whole lot of you! No bargains today. There is no other answer. You've asked me that before. My reply has always been the same, and it will always be the same. Now, get out!

CONTEMPT

Simms's boss has been inviting Simms and his wife to take a week's vacation with him. Simms is contemptful of the boss's tasts and intelligence, but the boss doesn't seem to notice and refuses to take, no, for an answer. Simms has just privately concluded arrangements to take another job, when he is called into the boss's office again.

SIMMS: I suppose you've asked me in again about that little vacation treat you've been offering me and Mrs. Simms--the reward for the extra contracts I've gotten for you--a week's trip with the Big Boss. It must seem strange to you--Sir--that I've been turning down your offer. Even if the trip was all planned nicely for your tastes, wasn't exactly what I might choose for myself, how could an employee turn you down? How could my wife and I not jump at the chance to share the luxuries of the boss's vacation: all the fancy things that appeal to a man with more cash than brains and more snobbery than good taste? I'll bet that's hard for you to understand. As a matter of fact, I'll bet a lot of things are. Well, my answer is very simple. You see, I can afford to make it clear and simple now. No. There is no other answer. You've asked me that before. My reply has always been the same, and it will always be the same. But, then, you're not likely to have occasion to ask again. You see, I've managed to find more rewarding and stimulating employment elsewhere, so I won't have to answer anymore of your foolish questions at all.

GRIEF

A young man is urging a friend, whose father died some time ago, to go back to school and take an interest in things again. The friend replies.

THE FRIEND: You want me to forget him--to forget or anyway half ignore my father's death. But don't you see that I can't? I know you mean it for my own good. I know that you want me to return to my work and my studies so that I can feel alive again. But life just seems to be meaningless now. He was my model and my guide for living. He came so near to showing me how to be truly and fully a man, and then he died before I could feel really grown and whole. Even then, when he was alive, I felt full and real when I was near him. And I loved and admired him more than any other person in the world. Now he's gone. There's no use your asking me to stop grieving and start living agin myself. There's no life left. There is no other answer. You've asked me that before. My reply has always been the same, and it will always be the same. I know you mean well, but no one can understand how I feel.

INDIFFERENCE

A man is speaking to a door-to-door solicitor for funds for the Anti-Vivisection Society.

THE MAN: Excuse me, Miss, but I'm not interested. I realize that you're collecting for what seems to you a worthy cause, and I'm sure that many of my neighbors around here would feel it is, too. But it's simply something that doesn't matter to me. You really might do much better at some of the other homes here. It's probably just that I'm not awfully sympathetic to animals. I see, in the abstract, that your society *has fine and touching ideals, but I don't feel really involved* enough to want to contribute funds. I appreciate that you've taken a lot of time here to try to show me what you're working for and to help me see if I might not be interested. But--though I've got nothing else to do this morning--I'm really afraid it's been a waste of your time. There is no other answer. You've asked me that before. My reply has always been the same, and it will always be the same. My own funds are pretty limited, and I save my contributions for the things that matter more to me personally. Thank you, though, for stopping by.

LOVE

A young man's father has threatened to cut him off without any money if he marries the girl of his choice. Although the girl herself is willing to try to manage anyway, she is afraid that the young man is not. He reassures her.

THE MAN: Darling, it doesn't matter what he does. My father can't control our lives. His money doesn't mean that much to me. You've assured me that you're willing to accept the life we can have together without it. Can you doubt, Dearest, that I'm willing to accept it, *too? I'd be willing to live far worse than that to have you at my* side to make living worthwhile. Of course I still want you! No matter what! Don't worry, Darling. I know it's only for my own sake that you even question it. But you just be sure of how much I love you, and never be afraid of anything. I love you. There is no other answer. You've asked me that before. My reply has always been the same, and it will always be the same. No one can change that or drive us apart.

APPENDIX B

Japanese Test Passage

English transliteration:

Mo hakani kotae yo wa naidesuyo

Onazi koto o nankai mo kikaretatte watashi no kotae

wa itsumo onazi desuyo.

Korekaramo kawarimasen.

Phonetic transcription:

moː hokáni kotaɛyoː ɔa náⁱndesɯyo?

Onáji kotó o nankai mo kikãrɛ taʔtt watas̆i nó kotaɛ

ɯa iʦumo onají désuyo.

Korɛ́karamo kawãri ɯa ãrimasén

(The English translation is the approximate equivalent of the English test passage, but personal pronouns were removed in order to fit the various contexts in Japanese without changing the wording for different relationships: "There is no other answer. This was asked before. The answer has always been the same, and it will always be the same.")

REFERENCES

Black, J. W., & Dreher, J. J. *Non-verbal messages in voice communication*. USN Sch. Aviat. Med. Res. Rep., 1955, No. NM001 104 500, 45.

Brody, M. W. Neurotic manifestations of the voice. *Psychoanal. Quart*., 1943, 12, 371-380.

Cohen, A. Estimating the degree of schizophrenic pathology from recorded interview samples. *J. clinic.Psychol*., 1961, 4, 403-406.

Cohen, A., & Starkweather, J. Vocal cues to language identification. *Amer. J. Psychol*., 1961, 74, 90-93.

Davitz, J. R., & Davitz, Lois J. The communication of feeling by content-free speech. *J. Communication*, 1959, 9, 6-13. (a)

Davitz, J. R., & Davitz, Lois J. Correlates of accuracy in the communication of feelings, *J. Communication,* 1959, 9, 110-117. (b)

Dibner, A. S. Cue-counting: a measure of anxiety in interviews. *J. consult. Psychol*., 1956, 20, 475-478.

Dittman, A. T., & Wynne, L. C. Linguistic techniques and the analysis of emotionality in interviews. *J. abnorm. soc. Psychol*., 1961, 63, 201-204.

Dusenbury, D., & Knower, F. H. Experimental studies of the symbolism of action and voice: II. A study of the specificity of meaning in abstract tonal symbols. *Quart. J. Speech*, 1939, 25, 67-75.

Eldred, S. H., & Price, D. B. A linguistic evaluation of feeling states in psychotherapy. *Psychiatry*, 1958, 21, 115-121.

Fairbanks, G. Recent experimental investigations of vocal pitch in voice. *J. acoust. Soc. Amer*., 1940, 11, 457-466.

Fairbanks, G., & Pronovost, W. An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monogr*., 1939, 6, 87-104.

Fairbanks, G., & Hoaglin, L. W. An experimental study of the durational characteristics of the voice during the expression of emotions. *Speech Monogr*., 1941, 8, 85-90.

Fletcher, H. *Speech and hearing in communication*. New York: D. Van Nostrand, 1953.

French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. *J. acoust. Soc. Amer*., 1947, 19, 90-119.

Goldman-Eisler, Frieda. Speech-breathing activity--a measure of tension and affect during interviews. *Brit. J. Psychol*., 1955, 46, 53-63.

Goldman-Eisler, Frieda, Speech-beathing activity and content in psychiatric interviews. *Brit. J. med. Psychol*., 1956, 29, 35-48. (a)

Goldman-Eisler, Frieda. The determinants of the rate of speech output and their mutual relations. J. psychosom. Res., 1956, 1, 137-143. (b)

Hargreaves, W. A., & Starkweather, J. A. Vocal behavior: an illunstrative case study. Paper read at Western Psychological Association, Seattle, June, 1961.

Kauffman, P. E. An investigation of some psychological stimulus properties of speech behavior. Unpublished Ph.D. dissertation, Univ. of Chicago, 1954.

Knower, F. H. Analysis of some experimental variations of simulated vocal expressions. J. soc. Psychol., 1941, 14, 369-372.

Kramer, E. The judgment of personal characteristics and emotions from non-verbal properties of voice. Psychol. Bull., in press. (a)

Kramer, E. Personality stereotypes in voice: a reconsideration of the data. J. soc. Psychol., in press, (b)

Krause, M. S. Anxiety in verbal bahavior: an intercorrelational study. J. consult. Psychol., 1961, 25, 272.

Krause, M. S., & Philisuk, M. Anxiety in verbal behavior: a validation study. J. consult. Psychol., 1961, 25, 414-419.

Mahl, G. F. Disturbances and silences in the patient's speech in psychotherapy. J. abnorm. soc. Psychol., 1956, 53, 1-15.

Mahl, G. F. Measuring the patient's anxiety during interview from "expressive" aspects of his speech. Trans. N. Y. Acad. Sci., 1959, 21, 249-257.

McQuown, N. A. Linguistic transcription and specification of psychiatric interview materials. Psychiatry, 1957, 20, 79-86.

Miron, M. E. A cross-linguistic investigation of phonetic symbolism. J. abnorm. soc. Psychol., 1961, 62, 623-630.

Ochai, Y., & Fukumura, T. On the fundamental qualities of speech in communication. J. acoust. Soc. Amer., 1957, 29, 392-393.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. The measurement of meaing. Urbana: Univ. of Illinois Press, 1957.

Pfaff, P. L. An experimental study of communication of feeling without contextual material. Speech Monogr., 1954, 21, 155-156.

Pittenger, R. E., & Smith, H. L. A basis for some contributions of linguistics to psychiatry. Psychiatry, 1957, 20, 61-78.

Pittenger, R. E., Hockett, C., & Danehy, J. The first five minutes. Ithaca, N. Y.: Martineau, 1960.

Pollack, I., Rubenstein, H., & Horowitz, A. Recognition of "verbal expression" over communications systems. Paper delivered at the 59th meeting of the Acoustical Society of America, 1960. (a)

Pollack, I., Rubenstein, H., & Horowitz, A. Communication of verbal modes of expression. *Language and Speech*, 1960, 3, 121-130. (b)

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Skinner, E. R. A calibrated recording and analysis of the pitch, force, and quality of vocal tones expressing happiness and sadness. *Speech Monogr*., 1935, 2, 81-137.

Soskin, W., & Kauffman, P. Judgment of emotion in word-free voice samples. *J. Communication*, 1961, 11, 73-80.

Starkweather, J. A. The communication value of content-free speech. Paper read at Midwest. Psychol. Assoc., Chicago, April, 1955. (a)

Starkweather, J. A. Judgments of content-free speech as related to some aspects of personality. Unpublished Ph.D. dissertation, Northwestern Univ., 1955, (b)

Starkweather, J. A. The communication-value of content-free speech. *Amer. J. Psychol*., 1956, 69, 121-123. (a)

Starkweather, J. A. Content-free speech as a source of information about the speaker. *J. abnorm. soc. Psychol*., 1956, 52, 394-402. (b)

Starkweather, J. A. Vocal communication of personality and human feelings. *J. Communication*, 1961, 11, 63-72.

Thompson, C. W., & Bradway, Katharine. The teaching of psychotherapy through content-free interviews. *J. consult. Psychol*., 1950, 14, 321-323.

Trager, G. L. Paralanguage: a first approximation. *Studies in Linguistics*, 1958, 13, 1-12.

RECOGNITION OF PORTRAYED EMOTION IN A FOREIGN LANGUAGE

Marijana Benesh, Ernest Kramer, and Harlan Lane
The University of Michigan

Recognition of Portrayed Emotion in a Foreign Language

Marijana Benesh, Ernest Kramer, and Harlan Lane
Behavior Analysis Laboratory
University of Michigan

In a recent review of research on the judgment of emotion from the nonverbal properties of speech, Kramer (1963a) concludes, "there is support for the popular notion that one can tell something about a person and his feelings from the sound of his voice..." All of the studies reported employed speakers to portray emotions and listeners to judge them who were from the same linguistic community.

Subsequently, Kramer (1963b) used Japanese speakers to portray emotions and found that American listeners were able to identify these emotions with considerable accuracy. We may infer that affect is coded similarly by speakers of quite diverse language and cultural backgrounds. The present study replicates this finding and examines the complementary case: Japanese listeners attempt to identify emotions portrayed by American speakers.

Method

Subjects

Three of the six speakers who portrayed the emotions were native American, male college students with some training in dramatics. The remaining three were native Japanese, male students, who had been in the United States less than six months and had a limited command of English.

Twenty-five American and nine Japanese students, of both sexes, comprised the two groups of listeners. They served in a single, half-hour session, conducted in a classroom.

Recordings of portrayed emotions

A common passage was chosen to represent each of the four emotions portrayed in the experiment. This technique for controlling differential cues to affect provided by the verbal "content" was first reported by Fairbanks (1940) and, later, by Kramer (1963b). The passage read:

"There is no other answer. You've asked me that before. My reply had always been the same, and it will always be the same."

The test passage was embedded in four different scripts, each of which described a situation appropriate to anger, grief, love, or indifference. The situations were common to both American and Japanese culture and the scripts were prepared in the language of the speaker. Each speaker read each script three times, while tape recordings were made in a radio studio. The 72 test passages obtained were extracted from context in random order, copied onto a second tape recording, and presented to listeners with high-fidelity equipment.

Procedure

Each subject was given a booklet containing 72 rating forms. Each form contained the typewritten words: love, indifference, grief, anger (or the corresponding terms in Japanese). Kramer (1963b) has shown that the order of listing the emotions on the rating form does not affect the frequency with which an emotion is identified. The listeners were asked to check, after each passage, the emotion they believed was portrayed by the passage.

Results and Discussion

Table 1 shows the per cent of affect ratings correctly assigned by Japanese and American listeners to each of the four emotions portrayed by the Japanese and American speakers. Note first the listeners' high accuracy in identifying portrayed emotion, unaided by cues from verbal "content", gesture, and the environmental context. On the average, subjects were 64% correct in recognizing the portrayed emotion; Kramer (1963b) found 67% correct recognition for the same emotions.

The data confirm the prior finding that American listeners can recognize emotions portrayed by Japanese speakers, and show that the converse is also true: Japanese listeners far exceed chance levels of accuracy in recognizing affect portrayed by American speakers. Evidently, speakers of two disparate languages code certain portrayed emotions in similar ways, since listeners from either language community can recognize their affect accurately. Dissimilarities in the coding of affect should also be noted, however. The Japanese listeners had difficulty in recognizing the portrayals of love, especially when rendered by Japanese speakers; English listeners were more successful, but were less accurate in their recognition of this emotion than of any other. Kramer (1963b) also reports low recognition scores for American listeners when love was portrayed by Japanese speakers.

It is interesting to note that English listeners proved more accurate in judging affect portrayed by the Japanese speakers than that portrayed by speakers of their own language; their scores were

68% and 79% correct, respectively (t for matched pairs = 5.9, $p < .01$). Furthermore, they were more often correct than the Japanese Ss in recognizing affect, not only in their own language ($t = 7.2$, $p < .01$) but also in Japanese ($t = 3.8$, $p < .01$).

Japanese listeners are more accurate in judging affect in their own language than in English (t for matched pairs = 3.30, $p < .05$). From these findings it follows that the Japanese speakers were more successful than the American in conveying the emotions they were instructed to portray. This outcome may be influenced by the fact that the Japanese speakers and listeners did have some contact with American language and culture prior to the experiment. In any event, the higher recognition scores of the Japanese speakers may not be safely attributed to their language or cultural background without a larger sample of speakers than that employed in this study.

Kramer (1963b) reports that American listeners who rank high in accuracy in judging affect in their own language tend to do likewise for a second language (Spearman rho = .41, $p < .05$). In the present study, correlations of the same form and magnitude were obtained for both the American listeners (rho = .38, $p < .05$) and the Japanese listeners (rho = .41, $p > .05$).

Summary

American and Japanese listeners identified with high accuracy the emotions: anger, grief, love and indifference, portrayed by American and Japanese speakers. Correct recognition of affect was most often associated with the English listeners and the Japanese speakers. Certain portrayed emotions are coded in similar ways by speakers of disparate languages.

References

Kramer, E. The judgment of personal characteristics and emotions from the nonverbal properties of speech. Psychol. Bull., 1963, (in press). (a)

Kramer, E. Judgment of portrayed emotion from normal English, filtered English, and Japanese speech. Unpublished doctoral dissertation, Univer. of Michigan, 1963. (b)

Fairbanks, G. Recent experimental investigations of vocal pitch in voice. J. Acoust. Soc. Amer., 1940, 11, 457-466.

Table 1

Per cent of affect ratings correctly assigned by 9 Japanese and 25 American listeners to four emotions, portrayed 3 times each, by 3 Japanese and 3 American speakers.

	Emotion Portrayed			
	Anger	Grief	Love	Indifference
English speakers English listeners	82	62	56	76
English speakers Japanese listeners	75	61	21	61
Japanese speakers Japanese listeners	90	84	15	56
Japanese speakers English listeners	83	93	61	86

INITIAL SPECIFICATIONS FOR A SPEECH AUTO-INSTRUCTIONAL DEVICE

INITIAL SPECIFICATIONS FOR A SPEECH AUTO-INSTRUCTIONAL DEVICE[1]

I. Introduction

A speech auto-instructional device (hereinafter called SAID) will be constructed and assembled from existing components for the purpose of teaching the prosodic features of speech to human speakers. Basically, SAID will be an electronic and electromechanical device that will serve the following overall functions:

a. recording and playback of a sequence of model patterns for presentation to the subject.

b. electronic analysis of certain parameters of simple and complex acoustic signals from two sources (the recorded mode and the subject).

c. electronic comparison of these parameter sets in time.

d. display of parameter mismatch for instructional and research purposes.

II. General description of functions

a. Record and playback. SAID will contain a four channel tape recorder. One channel will store model patterns, each no longer than three minutes, that are recorded by the experimenter and played back to the subject. Each of the other three channels will store a DC voltage that represents a parameter extracted from the model pattern (see below).

b. Analysis. SAID shall operate in any one or combination of three modes:

Mode 1) Pitch. It will continuously extract the pitch level and/or pitch rate of change of a tape-recorded model pattern (source 1) and of a speaker (source 2).

Mode 2) Amplitude. It will continuously extract the average speech power contour of a model pattern and of a speaker.

Mode 3) Tempo. It will continuously extract the tempo of a model pattern and of a speaker.

c. Comparison. In each selected mode, SAID will compare the acoustic parameter of the subject's speech to that of the recorded model pattern and generate an error signal proportional to the error in cps (pitch level), decibels, (average speech power) or milliseconds (tempo).

d. Display. This error signal shall provide information (shall be displayed) in four forms:

1. a difference voltage (DC) proportional to the error as a function of time for graphic recording.
2. the integral of the differencevoltage as a function of time for graphic recording.
3. the integral of the difference voltage as a function of time for display to the subject.
4. a discrete visual signal to the subject when this error voltage exceeds a preset threshold.
5. when SAID is set to operate in any combination of two modes, or in all three modes combined, a sum voltage shall be available which is the sum of the error voltages in each mode. It shall be possible to provide either this sum voltage or any one or more of the four display forms.

III. Detailed description of functions

a. Record and playback. The tape recording unit in SAID will contain one audio frequency record-playback system and three DC recording channels. The audio system, from microphone to playback headset shall have excellent frequency response over the speech spectrum (±2 db 80-8,000 cps) for this language instruction function. It shall be possible to employ either line or microphone inputs; the record amplifier shall have gain control and VU monitoring; there shall be two line impedance outputs with gain control and VU monitoring for driving the subject's earphones and a remote audiomonitor. The audio playback also feeds into parameter extraction circuitry (see below). As the model pattern is played once through, the outputs of the pitch, tempo, and amplitude extractors are sent to channels 2, 3, and 4 of the recorder where these parameters are recorded. For this purpose, an analog tape record/reproduce system employing frequency modulation may be appropriate. This system shall be rack mounted and so arranged as part of the total device that it may be removed and used for other applications when SAID is not in use.

After the model pattern and its parameters have been recorded, and the tape rewound, the tape recorder operates in the following fashion. The subject may imitate the model pattern either concurrently or alternately. When SAID is set to operate concurrently with the subject, the tape recorder begins playback when a voice-operated relay is triggered by the subject. (By this means the two signals to be compared begin in synchrony). The parameter extractors are automatically switched to analyze the audio input from the subject's microphone. The output of each extractor is sent, along with the corresponding stimulus parameter (from the tape deck) to the comparator. Depending on the mode of operation selected (pitch, amplitude, tempo), suitable error signals are sent from the comparator to the display units. The subject's microphone shall be part of a boom-type headset to minimize amplitude fluctuations due to head movements. It shall be possible to monitor the subject's speech at a remote station. (See physical characteristics below).

When SAID is set to operate alternately with the subject, the audio channel is played to the subject for the duration of the model pattern; the tape deck then rewinds rapidly and does not start again until the subject starts voicing. This time through, the external audio circuit is silenced but recorded parameters are sent to the comparator along with the subject's speech parameters, as in concurrent operation.

To permit SAID to recognize the end of a model pattern, and then to rewind, an inaudible coding tone, generated by internal circuitry, may be laid down, along with the original recording of the model pattern, at the end of eabh pattern. When the tone is played back to internal circuitry, the tape deck rewinds until the last coding tone, all error displays are cleared, and the tape segment is presented again upon voicing by the subject. The rewind effect of the coding tone may be held off by external circuitry so that the tape advances to the next pattern. Contact closure by external circuitry (not provided by the subcontractor) should initiate this event.

b. Analysis

Mode 1) Pitch. The pitch extraction circuit to be employed is shown in Figure 1. Since this circuit is the product of considerable research in pitch extraction methods,[2] substantive changes should not be made without consultation. The filter units shall be the high-and-low pass sections of two Dytronics No. 722 filters. The input attenuator shall be Hewlett-Packard No. 350 BR and the output frequency meter Hewlett-Packard No. 500 BR. These three devices shall be rack mounted and incorporated in SAID in such a way that they may readily be removed for other applications when SAID is not in use. The voltage amplifier shall be an audio frequency amplifier with good frequency response, high signal-to-noise ratio, and better than 30 db gain with variable gain control. Similar specifications apply to the mixer. The overall circuit and assembly of components shall be designed so as to optimize signal-to-noise ratio. An inexpensive oscilloscope shall be provided to monitor the sinusoidal input to the frequency meter. The DC output voltage of the meter, which is proportional to the input fundamental frequency, may be used as the pitch parameter to be recorded.

Mode 2) Amplitude analysis. The average speech power circuit to be employed is shown in Fig. 2. This circuit rectifies the complex speech signal and introduces an amount of filtering sufficient to remove the ripple at the fundamental frequency but not so great as to obscure speech transients (such as plosives). The bandwidth is 32 cps and the integrating time 11 msec. Its output may be used as the amplitude parameter to be recorded.

Mode 3) Tempo analysis. Detailed specifications for tempo-extraction circuitry are not provided. The principle of operation is as follows: the input to the tempo extractors shall be the output of the average speech power circuitry. Each time a peak occurs in each of these slowly varying DC voltages, a pulse is generated. The spacing among the pulses is here defined as the tempo of the pattern. It may be converted to an analog voltage by means of an interval timer and a digital-to-analog converter.

c. Comparison. The comparators receive concurrently the recorded speech parameters and their corresponding parameters extracted in real time from the subject's imitation of the model and generate an error signal which is proportional to the mismatch between the two inputs. In the pitch mode of operation, the difference in amplitude between two DC voltages is the error signal and this shall be calibrated in cycles per second. In the amplitude mode of operation, it is desired to compare the relative amplitudes of the model pattern with those of the subject. This requirement implies establishing a common reference level for the model and imitation amplitudes, perhaps at the beginning of the speech stream. In the tempo mode of operation, the difference in amplitude between the two voltage-analogs of the spacing of peaks in the speech power analysis is the error signal and this shall be calibrated in msec. When SAID is set to operate concurrently with the subject and the audio playback is fed into the subject's microphone, then the error signal in each mode must be zero. Furthermore, all amplifiers employed in the comparators (and display circuitry) shall have low drift. All equipment shall be operated from a highly stable power supply.

d. Display. The four forms of display are enumerated in section II d. The integral of the error voltage as a function of time, displayed to the subject, should be presented on a large and readily viewed panel meter. It shall be possible to change continuously the scale factor for the meter by known amounts so as to magnify or reduce the error voltage reading as necessary. The visual signal to the subject (a light on) is presented when the error voltage exceeds a preset threshold. Five threshold levels shall be preselectable by external circuitry, corresponding to the 20, 40, 60, 80 and 100% excursion points of the meter. It shall be sufficient to select a threshold that an appropriate set of contacts are closed by external circuitry. A contact closure shall be available in parallel with the error light to "inform" external circuitry that the error threshold has been exceeded. This information may simply result in a count on an external device, or it may be sent to the coding-tone-hold-off circuit on SAID, so that the tape now advances to the next model pattern.

IV. Monitoring

It shall be possible for the experimenter to monitor the subject's speech and the model pattern concurrently. For this purpose, an earphone jack, a loudspeaker, a variable gain control, and a selector switch to choose either output shall be associated with each source output (speaker and subject). A line impedance output shall also be available for tape recording. It shall be possible for the experimenter to monitor the output of all three comparators, whether or not the output is displayed on the subject's error meter. For this purpose, there shall be a meter for visual monitoring and a line impedance output for graphic level recording associated with each comparator output. A meter selector switch and a recording selector switch shall permit either the error (difference) voltage, or the integral of this voltage, to be monitored at the meter and/or recording outputs. It shall be possible for the experimenter to monitor the readings on the subject's error meter by means of a parallel meter located remotely. A parallel error light shall also be available, so that the experimenter may monitor the condition of the subject's light.

V. Physical characteristics

SAID shall consist of two major units: a subject's console and a control unit. The subject's console shall contain his error (panel) meter, his error light, receptacles for the cables that run to the control unit. Two ten-foot cables shall be provided: one funs from the console to a patch box; the other from the patch box to the control unit. Because electro-magnetic equipment may be used in the vicinity of SAID, the cabling should run between low impedance sources and loads. The control unit shall contain all the record, playback, extraction, logic, and monitoring circuitry of SAID, as well as the various selector switches, meters, graphic recording outputs, etc. It shall be possible to calibrate the error voltages provided for graphic recording in cycles per second for the pitch mode, decibels for the amplitude mode, and in milliseconds for the tempo mode.

The experimenter shall (1) select whether SAID operates concurrently with the subject or alternately with the subject, (2) select the mode, or dombination of modes, for which an error signal shall be displayed to the subject (seven-position switch), (3) select the error threshold that triggers the subject's error light and reports to external circuitry (five-position switch), (4) be able to reset the error meter at will, (5) be able to advance or rewind the tape at will.

Line impedance jacks (other than for cabling) shall be dual banana plug jacks; potentiometer controls shall have dial plates and shaft locks; relays shall be dust covered plug-in types; switching connections controlling or controlled by external circuitry shall terminate in NuWay studs on the experimenter's control panel.

Footnotes

1. Final specifications will be written in conjunction with the contractor; the device is scheduled for construction by September, 1963.

2. The pitch- and amplitude-extraction circuitry described in these specifications were developed at the Communications Sciences Laboratory, University of Michigan.

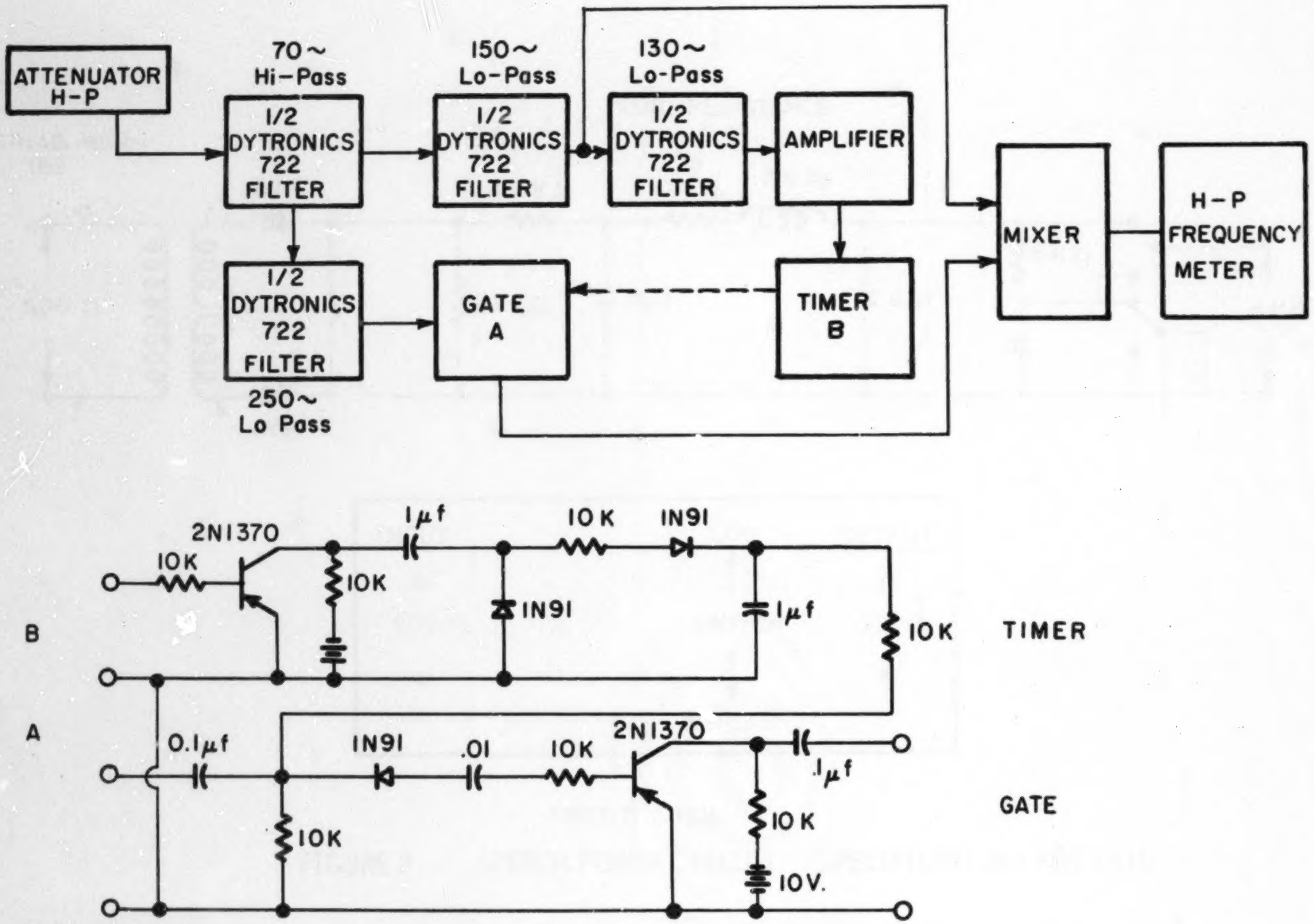
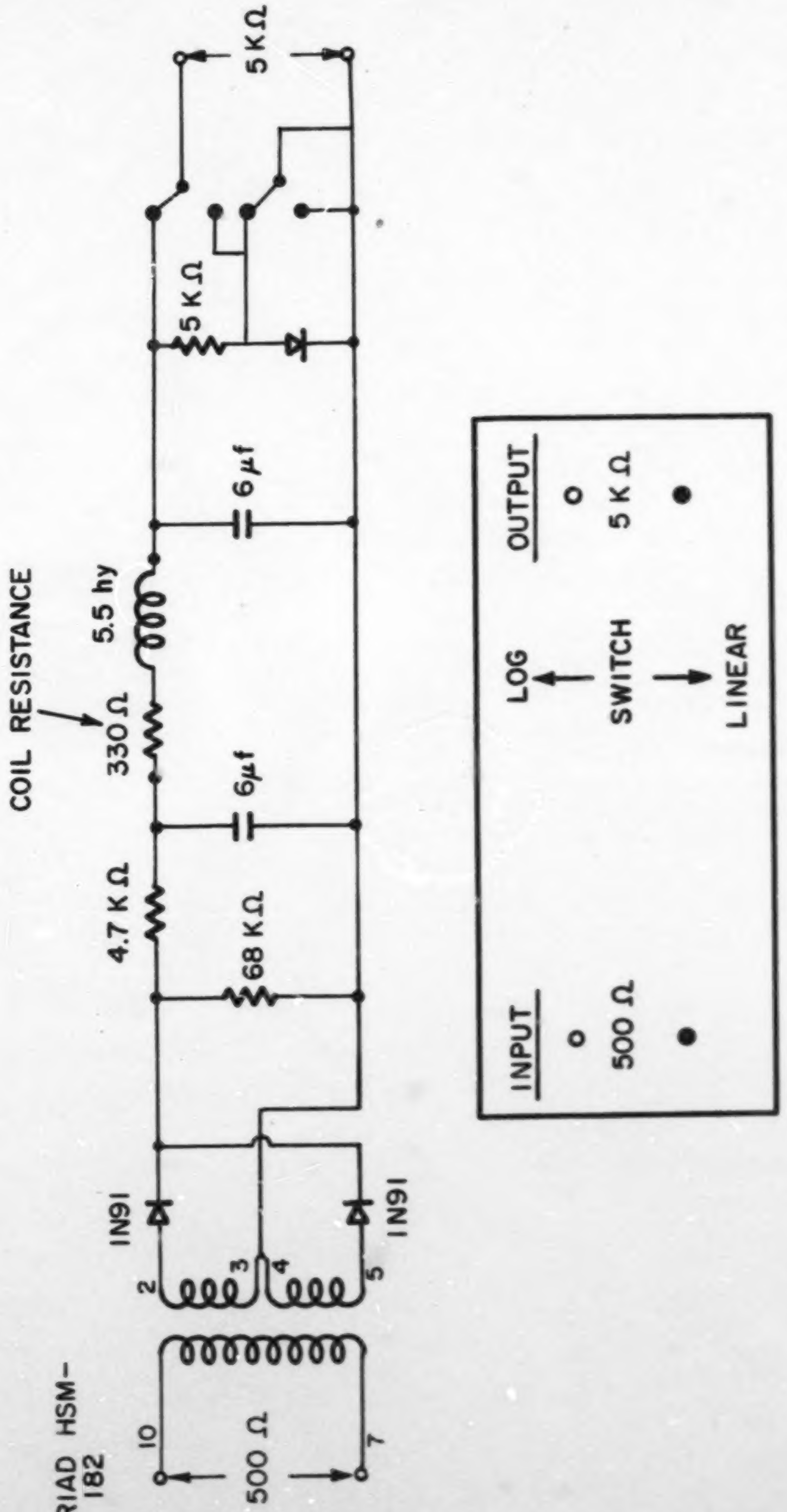

ATTENUATOR H-P
70~ Hi-Pass
1/2 DYTRONICS 722 FILTER
150~ Lo-Pass
1/2 DYTRONICS 722 FILTER
130~ Lo-Pass
1/2 DYTRONICS 722 FILTER
AMPLIFIER
MIXER
H-P FREQUENCY METER
1/2 DYTRONICS 722 FILTER
250~ Lo Pass
GATE A
TIMER B
2N1370
10K
10K
1μf
10K
IN91
IN91
1μf
10K
B
TIMER
A
0.1μf
IN91
.01
10K
2N1370
.1μf
10K
10K
10V.
GATE

FIGURE 1 PITCH EXTRACTOR (SAID SPECIFICATIONS)


TRIAD HSM-
182
10
7
500 Ω
2
3
4
5
IN91
IN91
4.7 K Ω
68 K Ω
6μf
COIL RESISTANCE
330 Ω
5.5 hy
6 μf
5 K Ω
5 K Ω
INPUT
500 Ω
LOG
SWITCH
LINEAR
OUTPUT
5 K Ω

FRONT VIEW

FIGURE 2 SPEECH POWER CIRCUIT SPECIFICATIONS FOR SAID

APPENDIX

STAFF

ARTICLES PUBLISHED

SPECIAL ACTIVITIES

Staff

Since March, 1962, the following personnel have been on the staff of the Behavior Analysis Laboratory.

H. L. Lane, Ph.D., Director
D. V. Cross, M.A., Research Assistant
E. F. Kramer, M.A., Research Assistant
(now at 321 South Anita, Los Angeles 49, California)
J. L. Kopp, M.A., Research Assistant
R. E. Skrocke, Research Engineer
D. R. Brinkman, Administrative Assistant
B. A. Schneider, Technical Assistant
T. M. Anderson, Technical Assistant
N. J. Kissock, Secretary
C. L. Hogue, Secretary
(now at 4633 Washtenaw, Ann Arbor, Michigan)

Articles published or accepted for publication[1]

Teaching machines and programmed learning. *Harvard Educ. Rev.*, 1961, 31, 469-476. (Harlan Lane)

Techniques of operant conditioning applied to second language learning. *Proc. XIV Int. Congr. Applied Psychol.*, 1961, Copenhagen. (Harlan Lane and F. R. Morton)

Experimentation in the language classroom: guidelines and suggested procedures for the classroom teacher. *Language Learning*, 1962, 12, 115-123. Also *U.S.I.A. English Teaching Forum*, 1963, 1, 3-8. (Harlan Lane)

Some differences between first and second language learning. *Language Learning*, 1962, 12, 1, 1-14. (Harlan Lane)

Reconditioning a consonant discrimination in an aphasic: an experimental case history. *J. Speech Hear. Disord.*, 27, 3, 1962, 232-243. (Harlan Lane and D. J. Moore)

On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization. *J. exp. Anal. Behav.*, 1962, 5, 4, 487-496. (D. V. Cross and Harlan Lane)

Methods and findings in an analysis of a vocal operant. *J. exp. Anal. Behav.*, 1963, 6, 179-188. (Harlan Lane)

Psychophysical parameters of vowel perception. *Psychol. Monogr.* (in press) (Harlan Lane)

Acquisition and transfer in auditory discrimination. *Amer. J. Psychol.* (accepted for publication). (Harlan Lane)

Repertoires in contact: studies of the interaction between first- and second-language learning. Ann Arbor: Publication of the University of Michigan Language Laboratory, Vol. VI, 1962, (in press). (Harlan Lane)

Foreign accent and speech distortion. _J. Acoust. Soc. Amer._ (in press). (Harlan Lane)

Methods for self-shaping echoic responding. _Modern Language J._ (in press) (Harlan Lane and Bruce Schneider)

Effects of training on estimates of vowel loudness. _J. exp. Psychol._ (accepted for publication) (Harlan Lane and D. V. Cross).

Personality stereotypes in voice: a reconsideration of the data. _J. Soc. Psychol._ (in press) (E. Kramer)

The judgment of personal characteristics and emotions from nonverbal properties of speech. _Psychol. Bull._ (in press) (E. Kramer)

Footnote

1. Reports of research conducted under U.S.O.E. contract 9265.

Special Activities

Addresses presented, since March, 1962, by Personnel of the Behavior Analysis Laboratory

1962 Northeast Conference on the Teaching of Foreign Languages. Boston, Massachusetts. April 13, 1962. "A new look at language learning". (Harlan Lane)

Eastern Psychological Association. Atlantic City, New Jersey. April 26-28, 1962. "On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization". (D. V. Cross)

Eastern Psychological Association. Atlantic City, New Jersey. April 26-28, 1962. "Methods for self-shaping echoic behavior". (B. A. Schneider)

Far Eastern Language-Area Liaison Conference. Chicago, Illinois. May 19, 1962. (Harlan Lane)

Acoustical Society of America. New York City, May 23-26, 1962. "Psychophysical parameters of vowel perception". (Harlan Lane)

Michigan Association of Mathematical, Physiological, and Experimental Psychologists, University of Michigan, Ann Arbor. June 4, 1962. "A relation between mechanics and acoustics in speech". (Harlan Lane)

Engineering Summer Conference, University of Michigan, Ann Arbor. June 27, 1962. "Learning theory and teaching machines". (Harlan Lane)

Psychonomic Society, St. Louis, Missouri, August 30, 31, September 1, 1962. "Influence of verbal conditioning on nonverbal behavior". (Harlan Lane)

Michigan Association of Mathematical, Physiological, and Experimental Psychologists, September 14, 1962. University of Michigan, Ann Arbor. "Control of auditory attention". (Harlan Lane)

American Association of College Governing Boards. Changing Functions of the College Teacher. University of Michigan, Ann Arbor. October 10, 1962. "Principles of programming". (Harlan Lane)

Acoustical Society of America, Seattle, Washington, November 7-10, 1962. 64th meeting. "Foreign accent and speech distortion". (Harlan Lane)

Michigan College Association 16th Annual Conference on Higher Education, University of Michigan, Ann Arbor. November 27-28, 1962. Theme: Programmed Learning, teaching machines, and similar auto instruction methods in colleges. "Current research in the new learning theory underlying programmed instruction". (Harlan Lane)

Phi Kappa Delta, Ann Arbor, Michigan, November 28, 1962. "Communication vs conditioning". (Harlan Lane)

Conference of Linguists and Psychologists, Stanford University, Stanford, California. March 7-8, 1963. (Harlan Lane)

University of Illinois, Urbana, Illinois. March 22, 1963. "Psychophysical determinants of speech perception and production". (Harlan Lane)

National Education Association Conference on Programmed Instruction, Washington D.C., March 24-26, 1963. "Programmed learning of a second language". (Harlan Lane)

Eastern Psychological Association, New York City, April 11-13, 1963. "Auditory generalization gradients of blind, retarded children". (Harlan Lane and Charles Curran)